

PIBIC / PIBITI



O LNCC convida:

XIII JORNADA DE INICIAÇÃO CIENTÍFICA E TECNOLÓGICA

Hall do auditório
17 de setembro de 2018
9h às 12h

Av. Getúlio Vargas 333. Petrópolis, RJ.



MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA,
INOVAÇÕES E COMUNICAÇÕES

Jornada de Iniciação Científica e Tecnológica do LNCC

Petrópolis, 17 de setembro de 2018.

Laboratório Nacional de Computação Científica – LNCC

Diretor
Augusto Cesar Gadelha

Coordenação de Gestão e Administração - COGEA
Anmily Paula dos Santos Martins

Coordenação de Métodos Matemáticos e Computacionais - COMAC
Frédéric Gerard Christian Valentin

Coordenação de Modelagem Computacional - COMOD
Márcio Arab Murad

Coordenação de Pós-Graduação e Aperfeiçoamento - COPGA
Abimael Fernando Dourado Loula

Coordenação de Tecnologia da Informação e Comunicação - COTIC
Wagner Vieira Léo

Programa Institucional de Bolsas de Iniciação Científica &
Programa Institucional de Bolsas de Iniciação em Desenvolvimento Tecnológico e Inovação
Marcos Garcia Todorov

Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq

Presidente
Mario Neto Borges

Coordenadora Geral do PIBIC/PIBITI
Lucimar Batista de Almeida

Jornada de Iniciação Científica e Tecnológica do LNCC

Comissão Interna do PIBIC/PIBITI-LNCC

Marcos Garcia Todorov
Eduardo Lucio Mendes Garcia
Helio José Correia Barbosa
Jack Baczynski

Avaliadores Externos

Alberto Torres Angonese - FAETERJ/UNIFESO
Eduardo Krempser da Silva - FIOCRUZ

Apresentação

O LNCC realiza este ano a XIII Edição da Jornada de Iniciação Científica e Tecnológica, que é um fórum de divulgação das pesquisas desenvolvidas no contexto dos Programas Institucionais de Bolsas de Iniciação Científica (PIBIC) e de Bolsas de Iniciação Tecnológica (PIBITI) fomentados pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). No período de agosto de 2017 a julho de 2018, o PIBIC e PIBITI congregaram alunos de várias instituições de ensino e de diversas áreas do conhecimento. Este volume apresenta os resumos dos trabalhos desenvolvidos pelos bolsistas no período. Durante a Jornada, os trabalhos são apresentados pelos bolsistas oralmente e em sessão de pôsteres e avaliados por um comitê científico externo e multidisciplinar.

Nesta XIII Edição da Jornada, o Comitê Externo de Avaliação do PIBIC/PIBITI tem a seguinte composição:

Prof. Alberto Torres Angonese - FAETERJ/UNIFESO

Prof. Eduardo Krempser da Silva - FIOCRUZ

Destacamos o papel relevante do PIBIC/PIBITI do LNCC no desenvolvimento das pesquisas no LNCC e, principalmente, na formação complementar dos bolsistas, promovendo o aprimoramento do conhecimento, espírito criativo, reflexão crítica e ética. Estas características têm contribuído para suas inserções no mercado de trabalho e em programas de pós-graduação, como o PPG em Modelagem Computacional do LNCC. Este é o resultado do esforço e dedicação de todos os participantes.

Agradecimentos

Agradecemos ao CNPq pelas bolsas concedidas, à Direção do LNCC pelo apoio e à Comissão Interna do PIBIC e PIBITI no LNCC.

Agradecemos a disponibilidade e contribuição dos membros do Comitê Externo de Avaliação. O sucesso desta Jornada, e do Programa como um todo, é o resultado da dedicação e do esforço de toda a comunidade do LNCC. Expressamos em particular nosso reconhecimento ao apoio concedido pela secretaria do PPG-LNCC e, em particular, à Sra. Roberta Machado.

Marcos Garcia Todorov
Coordenador do PIBIC/PIBITI - LNCC

Índice

Reposicionamento de Fármacos para a Doença de Chagas	01
Bolsista: Aldo Patrick Corrêa	
Orientador: Laurent Emmanuel Dardenne	
Coorientador: Isabella Alvim Guedes	
Uma Interface "Web" para o Procedimento de Decomposição em Ciclos de Redes Bioquímicas.....	08
Bolsista: Bruno Barreto Martins do Amaral	
Orientador: Marcelo Trindade dos Santos e Maurício Vieira Kritz	
Coorientador: Sicília Ferreira P. P. Júdice	
Modelagem de Incrustação de CaCO_3 em Trocadores de Calor	10
Bolsista: Carmen Lúcia Corrêa Bonifácio	
Orientador: Renato S. Silva	
Métodos Computacionais em Finanças: Utilização do Método de Monte Carlo na precificação da opção IDI.....	15
Bolsista: Diogo Bortolozo Henriques de Oliveira	
Orientador: Jack Baczynski	
Calibração de Modelos de Incrustação de CaCO_3 em Trocadores de Calor	17
Bolsista: Fernando Henriques Pereira Cardozo	
Orientador: Renato Simões	
Estudo do Mecanismo de Proteção de Células-Troncos Contra Apoptose induzido por drogas	23
Bolsista: Gabriel Moraes Kleiman	
Orientador: Regina Célia Cerqueira de Almeida	
Coorientador: Francisco J. P. Lopes	
Indexação Métrica em Junções Espaciais.....	33
Bolsista: Geovane Noel Alves	
Orientador: Fábio André Machado Porto	
Localização de Robôs Móveis via Filtro de Kalman.....	37
Bolsista: Hiago Riba Guedes	
Orientador: Marcelo Dutra Fragoso	
Métodos de Discretização para Sistemas Físicos Dispersivos e/ou Dissipativos.....	44
Bolsista: Igor P. dos Santos Pereira	
Orientador (a): Sandra M. Cardoso Malta	
Coorientador (a): Cristiane Oliveira de Faria	
O uso do Microsoft Kinect para manipular objetos 3D no CAVE	52
Bolsista: Jean Rosa Manhães	
Orientador: Jauvane C. de Oliveira	

Gerência e Análise de Dados Científicos de Workflows de Filogenia Aplicada a Genomas de Arbovírus.....	59
Bolsista: José Maria de Souza Junior	
Orientador (a): Kary Ann del Carmen Ocaña Gautherot	
Coorientador: Luiz Manoel Rocha Gadelha e André Elias Rodrigues Soares	
Modelagem e Controle de um Manipulador Robótico Sujeito a Falhas.....	64
Bolsista: Junior do Nascimento Xavier	
Orientador : Marcos Garcia Todorov	
Treinamento para Comissários de bordo.....	67
Bolsista: Leon Baldez Stelman	
Orientador: Jauvane C. de Oliveira	
Estudo e Aplicação de Técnicas de Visualização para o Problema da Progressão Tumoral.....	73
Bolsista: Matheus Evangelista de Oliveira	
Orientador (a): Regina C. Almeida	
Coorientador (a): Sicília Ferreira Ponce Pasini Judice	
Paralelização do Model-R baseada em Dataflow com Apache Spark.....	79
Bolsista: Matheus Machado da Rosa Albuquerque	
Orientador: Luiz Manoel Rocha Gadelha Junior	
Controle ótimo de vigas de Euler - Bermoulli.....	84
Bolsista: Monique Ribeiro da Costa	
Orientador: Jaime Edilberto Muñoz Rivera	
Avaliação de Métodos de Aprendizado em Aplicações de Saúde.....	88
Bolsista: Raquel de Abreu Junqueira Gritz	
Orientador: Fábio André Machado Porto	
Simulação Numérica e Computacional do Tráfego Viário.....	91
Bolsista: René Constancio Nunes de Lima	
Orientador: Elson M. Toledo	
Coorientador (a): Regina Célia P. Leal Toledo	
Métodos Numéricos para o escoamento Bifásico em Meios Porosos Heterogêneos em Ambientes Computacionais de Arquitetura de memória Híbrida.....	98
Bolsista: Weber Guilherme Dias Ribeiro	
Orientador (a): Carla Osthoff Barros	
Coorientador: Frederico Luis Cabral	

Bolsistas PIBITI

Inicialização de Fluidos para Animação Computacional.....	103
Bolsista: Allan Carlos Amaral Ribeiro	
Orientador: Gilson Antonio Giraldi	
Aplicação das Ferramentas Intel Parallel Studio para modernização de código para métodos numéricos de diferenças finitas para solução de equações diferenciais parciais em arquitetura Intel Haswel/Broadwell.....	109
Bolsista: Gabriel Pinheiro da Costa	
Orientador: Carla Osthoff	
Coorientador: Frederico Luis Cabral	
Técnicas de Aprendizagem de Máquina na Segmentação de Imagens Médicas de Ultrassom Intravascular	115
Bolsista: Jefferson da Silva Fernandes de Azevedo	
Orientador: Pablo Javier Blanco	
Processamento e Visualização de Imagens de Microscopia Confocal.....	117
Bolsista: Matheus Alves de Moura	
Orientador: Gilson Antonio Giraldi	

RELATÓRIO DAS ATIVIDADES REALIZADAS NO PROGRAMA
INSTITUCIONAL DE BOLSA DE INICIAÇÃO CIENTÍFICA

Reposicionamento de Fármacos para a Doença de Chagas

Período 08/2017 – 07/2018

Aldo Patrick A. Corrêa^{a,1}, Isabella A. Guedes^{b,1}, Laurent E. Dardenne^{c,1}

^a patrickredfield@gmail.com, (24) 22336179

^b isabella@lncc.br, (24) 22336121

^c dardenne@lncc.br, (24) 22336009, (24) 992072941

¹ Laboratório Nacional de Computação Científica, Avenida Getúlio Vargas, 333, Quitandinha, Petrópolis - Rio de Janeiro, CEP 25651-075 – Brasil.
posgrad@lncc.br, (24) 2233-6000

1. INTRODUÇÃO

O uso do termo "doenças negligenciadas" mostra-se comumente presente em relatos provenientes de países subdesenvolvidos e em desenvolvimento, referindo-se a quadros infecciosos que não possuem tratamento eficaz, financeiramente acessível e/ou de fácil administração. Os exemplos mais importantes desse grupo de doenças são Dengue, Leishmaniose, Malária, Doença de Chagas, Esquistossomose e Hanseníase. Todas estas doenças possuem expressiva prevalência no território brasileiro, atingindo uma parcela considerável da população. Mundialmente, estima-se que um sexto da população possua pelo menos uma destas doenças^{1,2}.

A doença de Chagas, causada pelo protozoário flagelado *Trypanosoma cruzi*, é uma das diversas enfermidades que correspondem a lista de doenças tropicais negligenciadas disponibilizada pela Organização Mundial da Saúde (OMS). Considerada uma das doenças parasitárias mais importantes da América Latina, a mesma afeta aproximadamente 7 milhões de pessoas no mundo inteiro, sendo responsável por aproximadamente 12.000 óbitos por ano³.

Na década de 60, os nitroderivados nifurtimox e benzonidazol mostraram-se os principais agentes na erradicação da infecção e, conseqüentemente, responsáveis por seu tratamento, sendo o benzonidazol utilizado até os dias de hoje. Entretanto, este fármaco possui atividade limitada, especialmente na fase tardia da infecção, além de possuir alta toxicidade e, conseqüentemente, ser responsável por diversos efeitos adversos e também por sua inefetividade contra algumas cepas resistentes⁴.

Por conta de tais fatores, a pesquisa financeiramente acessível, rápida e efetiva de novos fármacos contra a infecção mostra-se indispensável. A cada dia os esforços da comunidade científica no reforço do arsenal terapêutico contra a doença de Chagas, baseado em inibição enzimática, mostram-se mais e mais presentes.

A estratégia de reposicionamento de fármacos tem como objetivo identificar novos usos para fármacos já existentes, reduzindo significativamente o processo de validação destes compostos como novas terapias, uma vez que testes de toxicidade e tolerância já foram realizados previamente⁵. Diversos projetos têm sido desenvolvidos visando o reposicionamento de fármacos para várias doenças, inclusive doenças negligenciadas. Em um trabalho recente, Kaiser et al. avaliaram e confirmaram atividade antiprotozoária de 100 fármacos com potencial para o tratamento de doenças negligenciadas através de testes *in vitro*⁶. Apesar da confirmação da atividade antiparasitária destes fármacos, os alvos terapêuticos (*e.g.* enzimas, receptores) pelo qual estes fármacos exercem sua função não foi

descrita. No contexto do desenho racional de fármacos, é de grande importância identificar o alvo terapêutico pelo qual um composto exerce sua ação farmacológica, visando compreender as características estruturais que justifiquem sua potência e seletividade, guiando etapas posteriores de otimização molecular. Estudos de triagem virtual em larga baseiam-se na predição do modo de ligação e afinidade de pequenas moléculas no sítio de ligação do alvo terapêutico de interesse, sendo amplamente utilizados para identificar novos compostos promissores e buscar as moléculas receptoras de compostos conhecidamente ativos^{7,8}.

Desta forma, o presente trabalho envolveu a triagem virtual do banco de fármacos construído no trabalho de Kaiser et al. contra a enzima glicose-6-fosfato desidrogenase (G6PDH), considerada como um potencial alvo terapêutico do parasita *T. cruzi* para o tratamento da Doença de Chagas. A G6PDH é responsável pela catálise da oxidação da β -D-glicose-6-fosfato (G6P) em 6-fosfoglicono- δ -lactona, importante na via das pentoses-fosfato, sendo seu substrato responsável pela ocorrência de diversas reações imprescindíveis para a manutenção da vida do parasita⁹.

2. OBJETIVOS

O projeto visa a aplicação de técnicas de modelagem molecular para a exploração de potenciais alvos terapêuticos relacionados à Doença de Chagas, visando propor o uso de fármacos atualmente disponíveis comercialmente como possíveis compostos anti-tripanosomatídeos.

3. METODOLOGIA

O atual trabalho envolveu a seleção da enzima do *T. cruzi* descrita como potencial alvos terapêuticos para o tratamento da Doença de Chagas G6PDH⁹. A estrutura tridimensional foi coletada do banco de estruturas *Protein Data Bank* (PDB)¹⁰, sob o código 5AQ1⁹ (resolução de 2.65 Å, complexada com o substrato natural β -d-glicose-6-fosfato – G6P).

Utilizamos o banco de fármacos construído por Kaiser et al.⁶, composto de 100 fármacos atualmente aprovados pelo *Food and Drug Administration* (FDA) para o tratamento de diversas outras doenças, para buscar compostos com potencial para serem inibidores da enzima G6PDH para o tratamento da Doença de Chagas (reposicionamento de fármacos). A estrutura tridimensional dos compostos foi gerada utilizando o programa OpenBabel, e os átomos de hidrogênio adicionados pelo portal DockThor (www.dockthor.lncc.br).

Para validar a preparação das estruturas das enzimas e o protocolo de atracamento molecular utilizado, realizamos estudos de *redocking* dos compostos originalmente presentes no sítio de ligação das duas enzimas. Todos os experimentos de docking molecular deste trabalho foram realizados no portal de triagem virtual DockThor¹¹ (www.dockthor.lncc.br).

Os parâmetros da grade de energia foram definidos de acordo com o ligante de referência de cada enzima, utilizando caixa cúbica de 22 Å em cada dimensão e centro da grade X = 80, Y = 119 e Z = 60. Utilizamos os parâmetros padrão do algoritmo genético do programa DockThor (modo *Standard*): 1.000.000 de avaliações, 24 rodadas do algoritmo de busca com diferentes sementes e população de 750 indivíduos.

Os mesmos parâmetros da grade de energia e do algoritmo genético utilizados no *redocking* foram utilizados para os experimentos de triagem virtual dos compostos do banco de fármacos. A predição da afinidade de ligação foi feita com a função linear DockTScore, recentemente desenvolvida no nosso grupo de pesquisa GMMSB¹², disponível no portal DockThor.

4. RESULTADOS

De acordo com os resultados de *redocking* obtidos, consideramos que o protocolo de *docking* adotado foi considerado adequado, uma vez que foi possível encontrar a conformação experimental do ligante de referência como a solução de melhor energia predita pelo DockThor com RMSD igual a 0.435 Å (*i.e.* RMSD menor que 2.0 Å e considerado como sucesso) (**Error! Reference source not found.**).

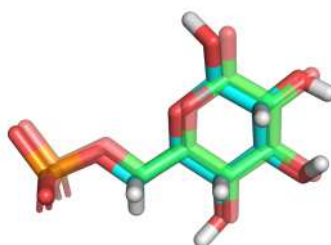


Figura 1. Conformação de melhor energia encontrada pelo programa DockThor (átomos de carbono em azul claro) sobreposta com a conformação observada experimentalmente para o ligante de referência presente na estrutura de código 5AQ1 – o substrato G6P (átomos de carbono em verde claro).

A triagem virtual com o banco de fármacos ativos contra *T. cruzi* encontrou 64 compostos com afinidade de ligação predita melhor que o ligante de referência (Score_{G6P} = -6.673 kcal/mol). Dentre estes, destacamos os 10 fármacos de melhor afinidade predita pela

triagem virtual na Tabela 1.

Tabela 1. Resultados da triagem virtual para os dez fármacos melhor ranqueados de acordo com a afinidade de ligação predita pela função DockTScore.

#	Fármaco	Função original	DockTScore ^a	Chagas IC ₅₀ (<i>in vitro</i>)
1	Lopinavir	Antiretroviral	-8.521	15.38 µM
2	Atazanavir	Antiretroviral	-8.391	33.34 µM
3	Itraconazole	Antifúngico	-8.001	0.004 µM
4	Cloperastine	Supressor de tosse	-7.950	9.46 µM
5	Rifabutin	Antibiótico	-7.767	27.51 µM
6	Ketoconazole	Antifúngico	-7.725	0.27 µM
7	Clofazimine	Antibiótico	-7.687	38.23 µM
8	Bifonazole	Antifúngico	-7.667	0.003 µM
9	Saquinavir	Antiviral	-7.758	17.14 µM
10	Triflupromazine	Antipsicótico	-7.566	16.71 µM

^a Afinidade de ligação predita em kcal/mol.

Todos os compostos foram preditos interagindo com o mesmo sítio de ligação do substrato G6P e realizando interações intermoleculares importantes com a enzima. O fármaco itraconazol foi predito como sendo o mais potente, interagindo através de uma ligação de hidrogênio com o resíduo Asp304 e outra com o cofator NADPH, além de interações lipofílicas com o sítio de ligação. Dentre os 10 fármacos preditos como sendo de melhor afinidade de ligação para a enzima G6PDH, três são significativamente potentes contra *T. cruzi*, com IC₅₀ menor que 1 µM e pertencentes à classe de antifúngicos: itraconazol, cetoconazol e bifonazol. Portanto, os resultados da triagem virtual indicam que a enzima G6PDH pode ser um possível alvo farmacológico destes fármacos, em que sua inibição pode estar relacionada à atividade tripanocida observada experimentalmente.

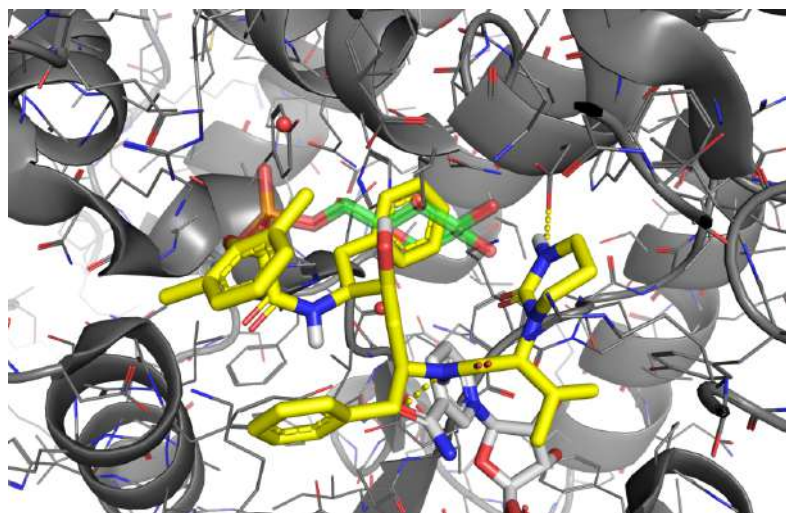


Figura 2. Modo de ligação predito para o fármaco itraconazol (amarelo) com o sítio de ligação da enzima G6PDH, sobreposto com o modo de ligação observado experimentalmente para o substrato natural G6P (verde).

5. CONCLUSÃO

Neste trabalho, foram utilizadas as estratégias de reposicionamento de fármacos e triagem virtual em larga escala para identificar fármacos com atividade tripanocida que pudessem ter como alvo terapêutico a enzima G6PDH do *T. cruzi*. O fármaco itraconazol foi predito como o composto mais potente contra a G6PDH, com afinidade predita significativamente melhor que do substrato natural G6P. Além disso, outros dois fármacos antifúngicos foram identificados como promissores inibidores da enzima G6PDH, sendo eles o cetoconazol e o bifonazol. Experimentos cinéticos futuros poderão validar a enzima G6PDH como alvo terapêutico destes fármacos tripanocidas.

6. REFERÊNCIAS

1. Yamey, G. & Torreele, E. The world's most neglected diseases. *BMJ* **325**, 176–177 (2002).
2. Neglected diseases: the strategies of the Brazilian Ministry of Health. *Rev. Saúde Pública* **44**, 200–202 (2010).
3. Sanchez-Sanchez, M., Rivera, G., A. Garcia, E. & Bocanegra-Garcia, V. Therapeutic Targets for the Development of Anti-Trypanosoma Cruzi Drugs: A Brief Review. (2016). Available at: <http://www.ingentaconnect.com/content/ben/mroc/2016/00000013/00000003/art00008#>. (Accessed: 21st July 2018)
4. Aldasoro, E. *et al.* What to expect and when: benznidazole toxicity in chronic Chagas' disease treatment. *J. Antimicrob. Chemother.* **73**, 1060–1067 (2018).
5. Ashburn, T. T. & Thor, K. B. Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* **3**, 673–683 (2004).
6. Kaiser, M., Mäser, P., Tadoori, L. P., Ioset, J.-R. & Brun, R. Antiprotozoal Activity Profiling of Approved Drugs: A Starting Point toward Drug Repositioning. *PLOS ONE* **10**, e0135556 (2015).
7. Dos Santos, R. N., Ferreira, L. G. & Andricopulo, A. D. Practices in Molecular Docking and Structure-Based Virtual Screening. *Methods Mol. Biol. Clifton NJ* **1762**, 31–50 (2018).
8. Guedes, I. A., de Magalhães, C. S. & Dardenne, L. E. Receptor–ligand molecular docking. *Biophys. Rev.* **6**, 75–87 (2014).
9. Mercaldi, G. F., Dawson, A., Hunter, W. N. & Cordeiro, A. T. The structure of a Trypanosoma cruzi glucose-6-phosphate dehydrogenase reveals differences from the mammalian enzyme. *FEBS Lett.* **590**, 2776–2786
10. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
11. de Magalhães, C. S., Almeida, D. M., Barbosa, H. J. C. & Dardenne, L. E. A dynamic niching genetic algorithm strategy for docking highly flexible ligands. *Inf. Sci.* **289**, 206–224 (2014).
12. Guedes, I. A., Barreto, A. M. S., Miteva, M. A. & Dardenne, L. E. Development of Empirical Scoring Functions for Predicting Protein-Ligand Binding Affinity. (Laboratório Nacional de Computação Científica, 2016).

Projeto PIBIC

Uma Interface “Web” para o Procedimento de Decomposição em Ciclos de Redes Bioquímicas

Aluno: Bruno Barreto Martins do Amaral

Orientadores: Marcelo Trindade dos Santos, Maurício Vieira Kritz, e Sicília Ferreira P. P. Júdice

Objetivos

Desenvolvimento de um portal ‘web’, que disponibilize para a comunidade científica algoritmos, procedimentos, e suporte computacional para a utilização do procedimento de decomposição em ciclos de redes metabólicas (bioquímicas) em escala genômica.

Introdução

Neste trabalho, estamos implementando um portal web que disponibilize o algoritmo de decomposição de redes metabólicas consistentes em massa desenvolvido, implementado e analisado (Vieira Kritz et al., 2010; Vieira Kritz e Trindade dos Santos, 2011) e testado em uma escala genômica em (Chapman et al., 2017). Nesta última referência ficou patente o grande impacto que o método de decomposição em ciclos pode ter no processo de investigação e compreensão das redes bioquímicas, e a importância de sua disponibilização para a comunidade científica.

Atividades Desenvolvidas

Estudo do procedimento de decomposição de redes metabólicas, entendendo o uso de grafos bipartidos e a aplicação do algoritmo de enumeração de Tarjan.

Estudo das linguagens de programação Python e C++. Estudo da linguagem de script Perl.

Estudo dos métodos de programação que implementam o procedimento de decomposição.

Estudo do script de decomposição em ciclos implementado na linguagem de programação Perl.

Elaboração da Interface com o Usuário (em andamento): Construção do wireframe das páginas que irão compor a interface web.

Participei como aluno regular do curso livre de cálculo e álgebra linear ministrado pelo Prof. Paulo César nas dependências do LNCC.

Resultados e Discussão

No período foram estudados os ciclos metabólicos, seus substratos e reações. A representação destes ciclos em grafos, a obtenção de ciclos destes grafos usando o algoritmo de Tarjan, e a implementação deste algoritmo na linguagem C++. Foram ainda analisados os dados de entrada e saída deste sistema segundo sua forma e conteúdo, assim como definidos os requisitos para a

implantação de um portal web que disponibilize a utilização deste procedimento para a comunidade científica.

Além disto, minha participação no curso livre de cálculo e álgebra linear foi essencial para me familiarizar com os conceitos e notações matemáticas usadas.

Conclusão

A importância deste trabalho se dá pela oportunidade de se disponibilizar o método de decomposição em ciclos como uma nova ferramenta para a investigação e compreensão das redes bioquímicas, para a comunidade científica.

Esta bolsa de iniciação científica foi meu primeiro contato com o trabalho de pesquisador e mudou minhas perspectivas a respeito de pós-graduação nas formas de mestrado e doutorado, me apresentando tanto às ferramentas e procedimentos adotados, assim como às expectativas e requerimentos da área.

Bibliografia

Vieira Kritz, M. et al., 2010. Organising metabolic networks: Cycles in flux distributions. *Journal of Theoretical Biology*, 265(3), pp.250–260.

Vieira Kritz, M. & Trindade dos Santos, M., 2011. Dynamics, Systems, Dynamical Systems and Interaction Graphs. In *Dynamics, Games and Science II*. Springer Proceedings in Mathematics. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 507–541.

Chapman, S.P. et al., 2017. Cyclic decomposition explains a photosynthetic down regulation for *Chlamydomonas reinhardtii*. *BioSystems*, 162, pp.119–127.

Python 3 Tutorial. SoloLearn. Disponível em: <<https://www.sololearn.com/Course/Python/>>. Acesso em: 23 jul. 2018.

The Comprehensive Perl Archive Network - www.cpan.org. Disponível em: <<https://www.cpan.org/>>. Acesso em: 23 jul. 2018.

1. Dados

Título: Modelagem de Incrustação de CaCO₃ em Trocadores de Calor

Bolsista: Carmen Lúcia Corrêa Bonifácio

Orientador: Renato S. Silva

Bolsa: PIBIC - CNPQ

Período avaliado: 01/04/18 a 06/07/2018

2. Objetivos

Implementar modelos matemáticos através de um programa computacional desenvolvido em linguagem C que apresente resultados com baixo desvio, ou seja, muito próximos aos obtidos em experimentos de bancada, aplicados inicialmente para trocadores de calor do tipo duplo tubo envolvendo diferentes cenários em relação as variáveis estudadas.

3. Introdução

O fenômeno de incrustação de CaCO₃ em trocadores de calor do tipo duplo tubo quando não controlado, afeta vários tipos de indústrias gerando parada de produção sem programação e elevados custos associados.

Na literatura existem diversos modelos como exemplo utiliza-se as referencias citadas por Bohnet [1]. Devido a essa grande quantidade, determinar o melhor modelo para cada situação exige um conhecimento maior sobre as variáveis analisadas e os fatores limitantes. A maneira usual da escolha de um modelo em detrimento de outros é comparar o desempenho dos modelos com um conjunto de dados similares ao da situação que se deseja a previsão. Assim podemos escolher o modelo com melhor desempenho em um determinado critério.

Entretanto devido a existência de uma grande variabilidade nas condições de operação e variáveis associadas a estes cenários, dificilmente um unico modelo será o ideal em todas as situações, por isso o estudo de um conjunto de modelos permite uma previsão mais ampla e global.

4 Material e Métodos ou Metodologia

Para o estudo dos modelos existentes de incrustações de CaCO₃ em trocadores de calor do tipo duplo tubo foi necessário compreender os as equações propostas por Mathias Bohnet [1].

Dos modelo existentes, este em estudo, parte da equação (1) na qual a taxa de massa de sólido depositada pro unidade de área variando em função do tempo é dada por:

$$\frac{dm_f}{dt} = \frac{dm_d}{dt} - \frac{dm_r}{dt} \quad (1)$$

onde m_d representa a taxa de crescimento devido a deposição e m_r o descréscimo da massa de sólido devido a remoção.

O Modelo Linear é representado pela equação (2).

$$\frac{dm_d}{dt} = B * (CF - Cs) \quad (2)$$

onde B é o coeficiente de transferência de massa e (CF – Cs) representa a difusão do sal na camada limite entre o fluido e o cristal formado.

O Modelo de Mathias Bohnet [1] parte da equação (1) e usa a equação (3) para expressar o resultado obtido em fator de Incrustação R_f .

$$m_f = Q_f * \lambda_f * R_f \quad (3)$$

Q_f = Densidade da camada de incrustação

λ_f = Condutividade térmica da camada de incrustação

Desta forma a equação proposta por Bonhet [1] é dada pela expressão (4).

$$\frac{dR_f}{dt} = \frac{1}{Q_f \lambda_f} * \frac{dm_f}{dt} \quad (4)$$

A partir deste conjunto de equações e dados sobre os componentes que afetam a deposição do sal conhecidos através da matriz de fases proposta pela literatura Epstein [2], as 5 fases pelas quais o fouling percorre desde a formação do primeiro cristal de sólido até a sua adesão à parede dos tubos do trocador, foi utilizado o método de Euler [3] que é baseado expansão da série de Taylor para obtenção dos resultados.

5. Resultados e Discussão

Os dois modelos estudados até a presente data, período de 3 meses de dedicação são: o Linear e o de Bohnet. Os calculos do fator de incrustação foram aplicados para um total de 300 horas de operação de um trocador de calor do tipo duplo tubo.

Utilizando linguagem C foi elaborado um programa cujos resultados foram plotados em gráficos usando conhecimentos da ferramenta GNUPLOT e apresentados abaixo.

Importante ressaltar que no experimento de Bohnet [1] o composto utilizado foi o Sulfato de Cálcio ($CaSO_4$), sal com comportamento semelhante ao carbonato de cálcio.

O modelo Linear, conforme previsto para esta situação, não representa adequadamente os pontos experimentais, isso porque não utiliza em suas variáveis de cálculo o componente da taxa de remoção de massa que neste caso de estudo é fator determinante.

Apresenta fator de incrustação com crescimento linear demonstrado pela reta na Figura 1.

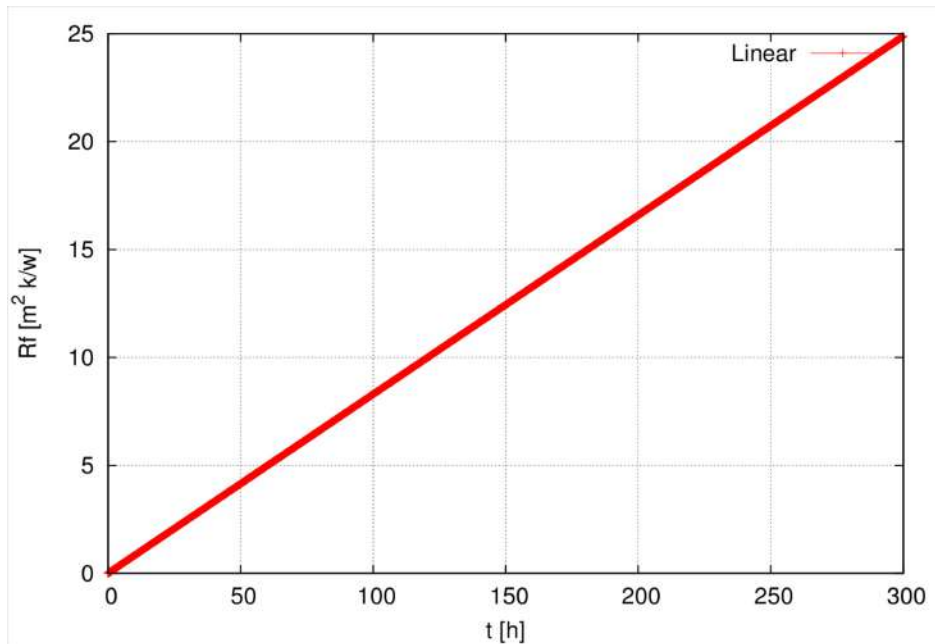


Figura 1. Modelo Linear: Fator de Incrustação por cristalização.

O modelo de Bohnet equação (4), calculado através do método de Euler, quando comparado com os valores obtidos experimentalmente se mostrou muito adequado. Apresentando curva assintota, comportamento confirmado devido aos dois componentes observados na equação (1). Figuras 2 e 3.

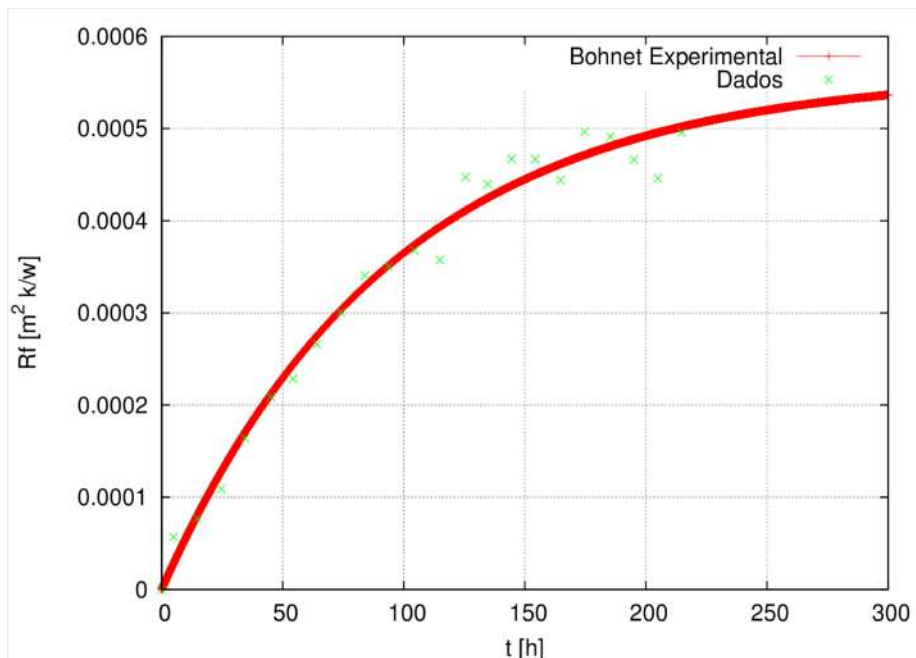


Figura 2. Fator de Incrustação usando valores de constantes medidas durante experimento para taxa de reação e fator de adesão em comparação com os dados experimentais.

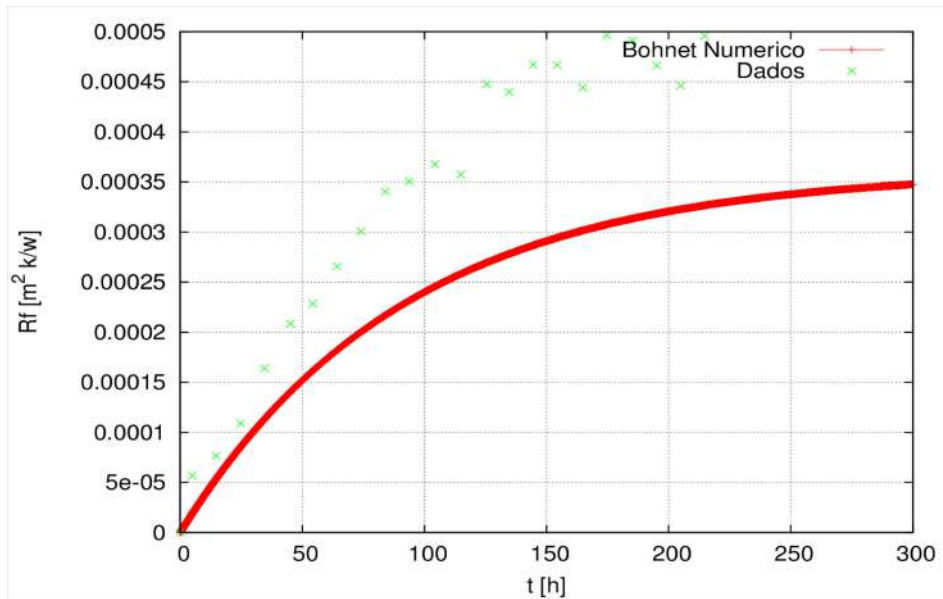


Figura 3. Fator de Incrustação usando somente valores numéricos em comparação com os dados experimentais.

Observa-se na Figura 3 um leve afastamento do fator de incrustação em relação aos dados experimentais, o que não acontece na Figura 2. Isso porque na Figura 2 o cálculo é realizado utilizando dois valores constantes obtidos por medição durante o teste de bancada de Bohnet [1]. Esses valores aproximam mais os resultados obtidos dos dados pontuais do teste.

A Figura 4 apresenta os dados calculados numericamente versus os dados calculados com o uso das constantes medidas no experimento plotados juntamente com os dados experimentais.

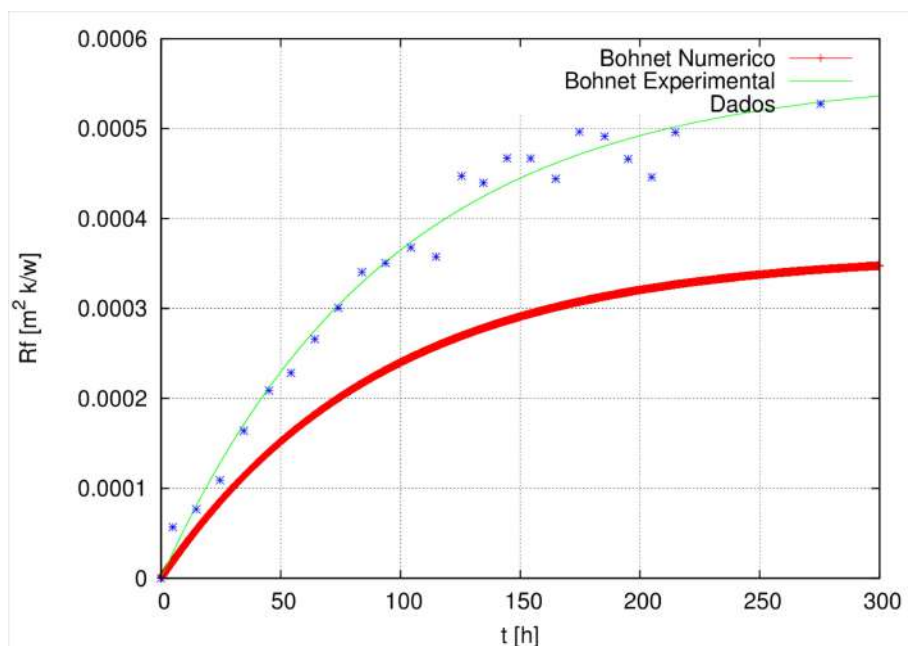


Figura 4. Fator de Incrustação Numérico x Fator de Incrustação do Experimento.

6. Conclusões

Considerando a gama de variáveis apresentadas na matriz de fases que afetam a formação de incrustação e com base nos estudos já realizados, para este caso em questão, o modelo numérico se apresentou adequado para ser aplicado na determinação de incrustação em trocadores de calor.

É necessário continuar a linha de estudos para ampliar a gama de modelos testados e comprovados.

A metodologia aplicada para desenvolvimento dos trabalhos se mostrou eficaz e conclusiva. Houve checagem dos resultados obtidos de fator de incrustação através do programa desenvolvido numericamente com os dados experimentais. Este procedimento funcionou como métrica de performance e mostra que este modelo teórico é satisfatório para prever incrustações e auxiliar indústrias de diversos segmentos a minimizar perdas e reduzir custos originários destas ocorrências indesejáveis no processo produtivo.

7. Referências Bibliográficas

[1] Mathias Bohnet. Fouling of Heat transfer Surfaces. Chemical Engineering Technology, 10(1987) 113-125.

[2] Norman Epstein. Thinking about heat transfer fouling: A5X5 Matrix. Department of Chemical Engineering, University of British Columbia, Vancouver, B.C, Canada, 2007.

[3] Gerald Rocktenwald. Numerical Integration of Ordinary Differential Equations for Initial Value Problems. Department of Mechanical Engineering, Portland State University, 2006.

RELATÓRIO FINAL DE INICIAÇÃO CIENTÍFICA

IDENTIFICAÇÃO

Título do Plano de trabalho:

Métodos Computacionais em Finanças: Utilização do Método de Monte Carlo na precificação da opção IDI

Nome do Bolsista: Diogo Bortolozo Henriques de Oliveira

Nome do Orientador: Jack Baczynski

Local de execução: Laboratório Nacional de Computação Científica (LNCC).

Vigência do plano de trabalho: 14/08/2017 a 14/08/2018.

INTRODUÇÃO

O presente trabalho tem por finalidade criar simulações que facilitem o entendimento do mercado financeiro, e que sejam possivelmente utilizadas para a aplicação real diminuindo riscos e otimizando resultados.

METODOLOGIA

Foram utilizados no trabalho livros para tomada de conhecimento do tema de precificação de derivativos financeiros. A seguir empregou-se o Software Matlab que auxilia na implementação dos resultados das simulações do preço da opção de venda sobre ações e de um Bond zero coupon em renda fixa [2].

DESENVOLVIMENTOS

Com base em métodos computacionais e matemáticos, utilizamos teoremas de Black Scholes para calcular o preço da opção e comparamos o valor com as simulações utilizando Monte Carlo. Utilizou-se o modelo de Black and Scholes: parâmetros constantes e movimento Browniano Geométrico. Chegamos num valor bem aproximado.

Assumimos que o preço X_t de um ativo de risco (stock ou ação) no tempo t é dado pelo movimento geométrico browniano, onde, como de costume $W = (W_t, t \geq 0)$, é o movimento browniano que podemos escrever formalmente como [3]:

$$dX_t = cX_t dt + \sigma X_t W_t \quad (1.1)$$

Feito a simulação, calculamos o valor da opção de venda, dita:

$$Call = E[e^{(-rt)} \max\{S_T - K, 0\}] \quad (1.3)$$

Dentro dos resultados encontrados, comparamos o valor com o resultado da formula de Black Scholes [1]:

$$D1 = (\log\left(\frac{s}{x}\right) + \left(r + \frac{v^2}{2}\right)t) / v\sqrt{t} \quad (1.4)$$

$$D2 = D1 - v\sqrt{t} \quad (1.5)$$

$$Call = SN(D1) - xe^{(-rt)}N(D2) \quad (1.6)$$

Está em curso a simulação do preço do Bond por Monte Carlo e da formula fechada utilizando o modelo de Vasicek [4]. No passo seguinte vamos evoluir o derivativo para o IDI.

RESULTADOS

Os valores da simulação com a formula ficaram bem aproximadas. Os resultados dependeram do fator processamento de máquina pois a simulação de monte Carlo é uma técnica baseada em amostras massivas de números aleatórios. Porém os resultados foram promissores. Conseguimos criar uma simulação que se aproxima da formula de Black and Scholes, podendo assim aplicar essas simulações para criar cenários futuros com mais exatidão. Os resultados em renda fixa estão em curso.

BIBLIOGRAFIA

- [1] Hull, J.C. Opções, Futuros e Outros Derivativos 9º edição.
- [2] Brandimarte, P. Numerical Methods in Finance and Economics 2º edition.
- [3] Mikosch, T. Elementary Stochastic Calculus, World Scientific.
- [4] Glasserman, P., Monte Carlo Methods in Financial Engineering, Springer.

Calibração de Modelos de Incrustação de $CaCO_3$ em Trocadores de Calor

Bolsista: Fernando Henrique Pereira Cardozo

Orientador: Renato Simões

Julho 2018

1 - Objetivo

O principal objetivo do trabalho é implementar modelos computacionais que usem combinação de modelos existentes na literatura de forma a melhorar a precisão das previsões. Esse objetivo passa pelo estudo da calibração de modelos conhecidos da literatura.

2 - Introdução

A incrustação de $CaCO_3$ em trocadores é um problema que afeta diversas indústrias e para entender e prever sua ocorrência se faz uso de modelos matemáticos e numéricos.

Como o objetivo é encontrar o modelo com o melhor resultado e sabendo que tem-se inúmeros modelos na literatura é necessário que sejam feitas comparações entre os resultados encontrados por cada modelo e os dados utilizados por cada um deles para que se possa fazer uma comparação.

3 - Material e Metodologia

No início desse trabalho foi estudado o método de Euler para soluções numéricas. O Método aproxima as soluções através da derivada da função, como segue abaixo,

$$y(t_i) = y(t_{i-1}) + h \left. \frac{dy(t)}{dt} \right|_{t=t_{i-1}} \quad (1)$$

O estudo do Método de Euler foi feito através de uma modelagem para a solução de um exemplo, que possui solução analítica de forma a se comparar a importância de uma boa

escolha do valor de h , que é o tamanho do intervalo onde se calculará cada derivada, segue o exemplo,

$$\frac{dy(t)}{dt} = t - 2y(t) \quad y(0) = 1 \quad (2)$$

A solução analítica da equação diferencial acima é dada por,

$$y(t) = \frac{1}{4}[2t - 1 + 5e^{-2t}] \quad (3)$$

Ao se concluir o trabalho para o método de Euler e entender a importância da escolha de um valor de h que nos retorne uma solução precisa o suficiente, foi dado início ao estudo da incrustação através do modelo utilizado no paper, *Fouling of Heat Transfer Surfaces - Matthias Bohmet*, onde ele trabalha uma modelagem para a incrustação na superfície de trocadores de calor, de forma a prevê-los.

Nosso objetivo é, utilizando a modelagem feita no paper, entender o processo de incrustação por cristalização e modelar computacionalmente uma aproximação numérica capaz de prever o processo.

O processo de deposição pode ser expresso pela deposição por unidade de área m_f ,

$$\frac{dm_f}{dt} = \dot{m}_d - \dot{m}_r \quad (4)$$

Com \dot{m}_d sendo a taxa de deposição e \dot{m}_r a taxa de remoção de material.

Foram realizados dois cálculos:

Um supondo \dot{m}_r nula, *modelo 1*, ou seja um crescimento ilimitado da deposição, utilizando a seguinte equação:

$$\left(\frac{dm_f}{dt}\right)_d = \dot{m}_d = \beta \left\{ \frac{1}{2} \left(\frac{\beta}{k_r}\right) + (c_F - c_S) - \sqrt{\frac{1}{4} \left(\frac{\beta}{k_r}\right)^2 + \left(\frac{\beta}{k_r}\right) (c_F - c_S)} \right\} \quad (5)$$

E o segundo utilizando tanto o termo de deposição, quanto o de remoção, sendo o de remoção, *modelo 2*:

$$\left(\frac{dm_f}{dt}\right)_r = \dot{m}_r = \frac{K_6}{P} \rho_f (1 + \delta \Delta T) d_p (\rho^2 \eta g)^{1/3} \frac{m_f}{\rho_f} w^2 \quad (6)$$

Como a grandeza de maior importância, e a qual temos parâmetros de comparação no paper, é o fator de incrustação (*fouling factor*) R_f que pode ser obtido a partir de m_f pela seguinte relação,

$$R_f = \frac{m_f}{\rho_f \lambda_f} \quad (7)$$

4 - Resultados

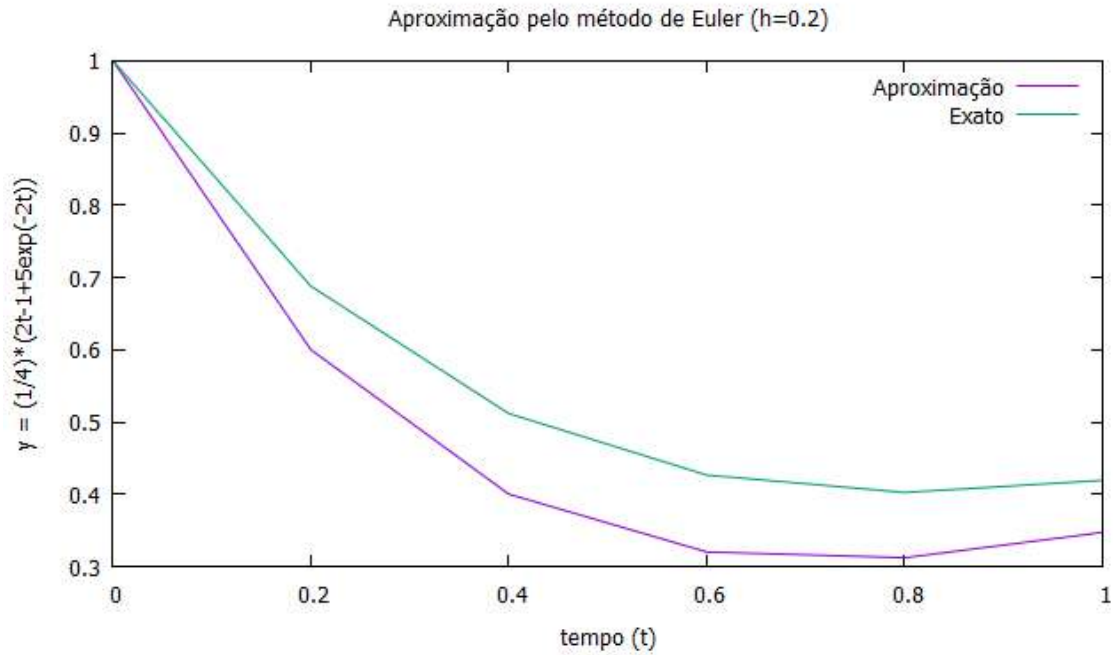


Figura 1: Solução da Equação 2, pelo método de Euler para $h = 0.2$

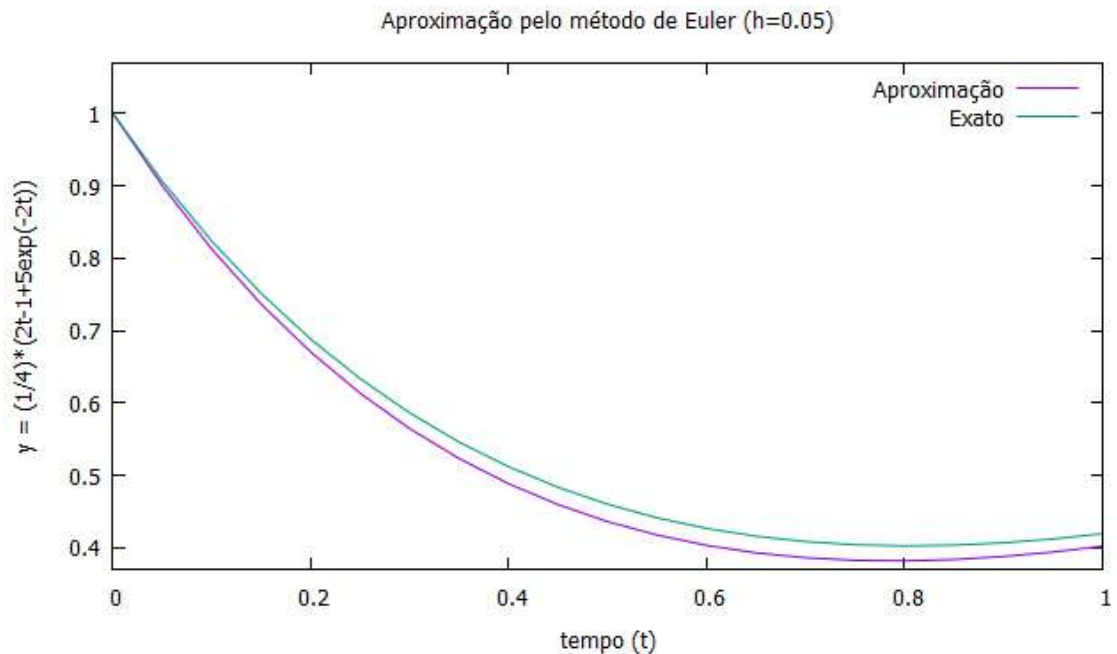


Figura 2: Solução da Equação 2, pelo método de Euler para $h = 0.05$

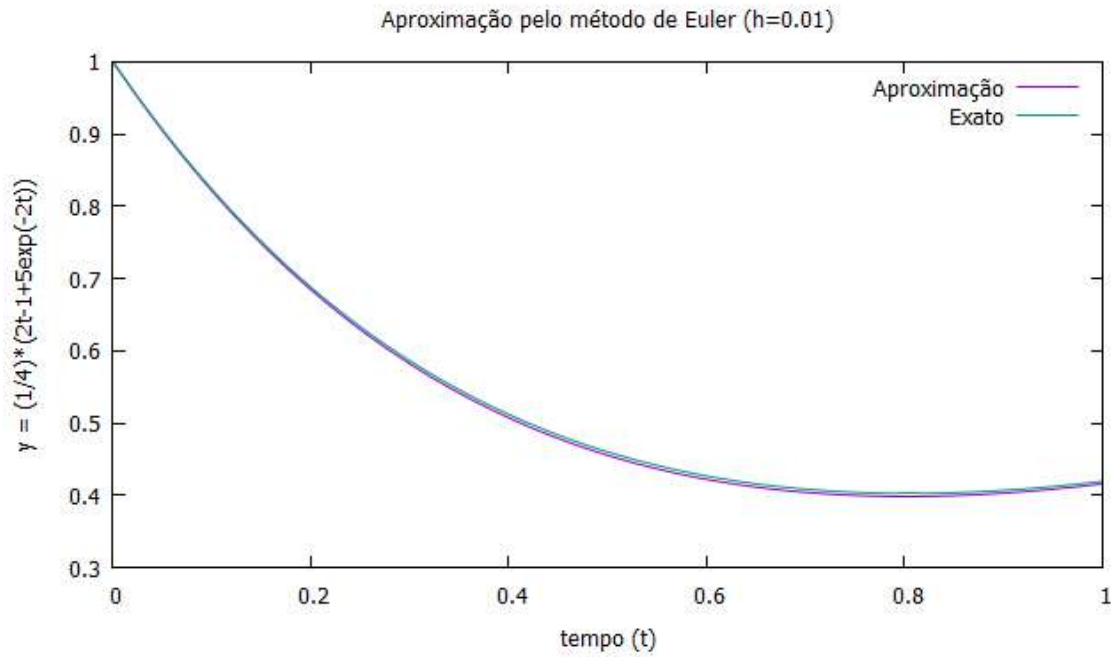


Figura 3: Solução da Equação 2, pelo método de Euler para $h = 0.01$

A partir da análise dos gráficos vemos que com uma diminuição do valor de h a precisão da aproximação numérica pelo método de Euler é maior. Quando utilizamos um $h = 0.01$ como mostrado na figura 3, praticamente não é mais possível ver a diferença entre a solução analítica e a solução numérica.

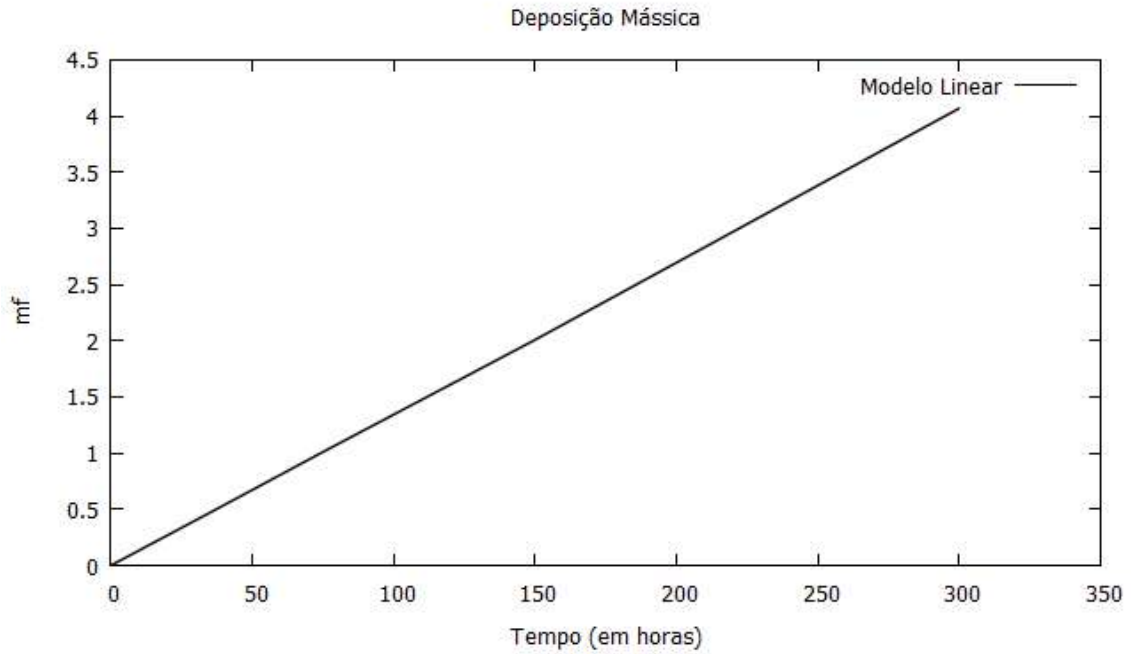


Figura 4: *Modelo 1* - Deposição linear.

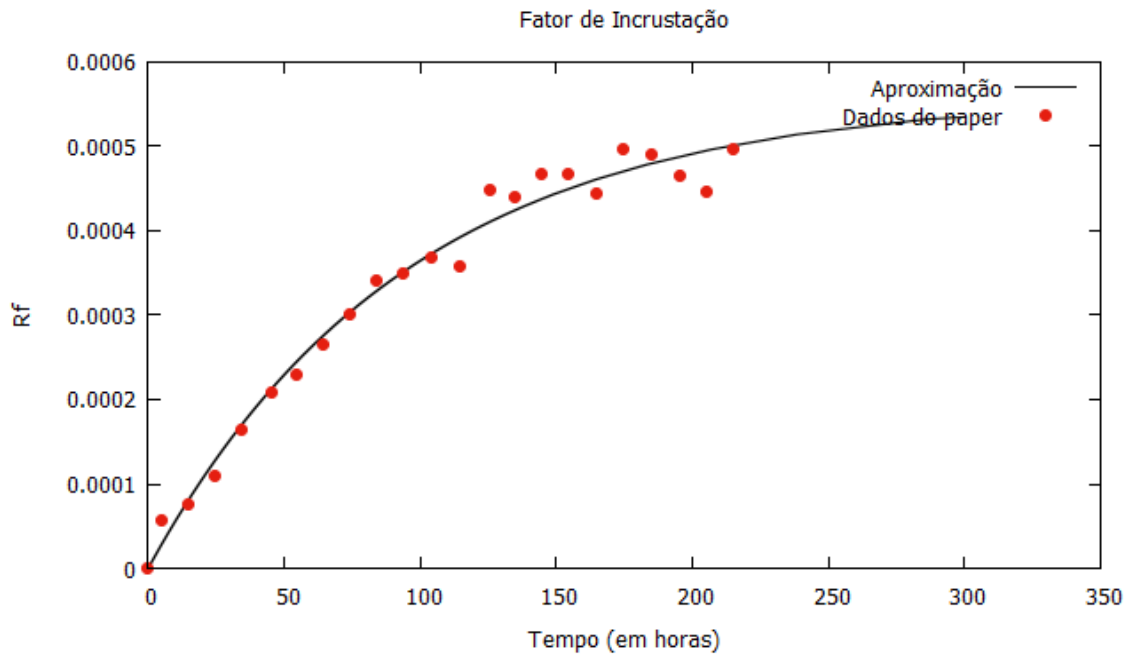


Figura 5: *Modelo 2* - Fator de Incrustação, calculado a partir das equações 4 e 7, com os pontos experimentais obtidos do paper.

A figura 4, nos mostra, como esperado que se houver apenas o termo de deposição (hipótese feita para o modelo 1), o crescimento será linear e ilimitado, porém

essa hipótese não demonstra corretamente a realidade, uma vez que associada à deposição, ocorre a remoção das camadas de incrustação.

Sendo assim foi trabalho outro modelo, onde considera-se os dois termos, como mostrado na figura 5, retratando assim, uma situação mais real.

Uma vez que se tinha em mãos os dados usados pelo Bohnet no paper, foram plotados esses pontos em sobreposição a curva encontrada utilizando-se o modelo 2 para observarmos o resultado.

5 - Conclusão

O primeiro trabalho, com o método de Euler tinha por objetivo mostrar a influência do h na qualidade do resultado de uma aproximação numérica e como esperado, verificamos que a precisão dos resultados cresce com a diminuição de h .

Ao se utilizar as hipóteses feitas através dos modelos 1 e 2 para incrustação, esperávamos encontrar resultados compatíveis com os dados experimentais dos quais se tinha posse e através de uma análise do gráfico na figura 5, podemos entender que o resultado é aceitável, uma vez que a curva representa bem os pontos.

6 - Referências Bibliográficas

Fouling of Heat Transfer Surfaces - Matthias Bohnet (1987)

Numerical Integration of Ordinary Differential Equations for Initial Value Problems - Gerald Recktenwald

**Laboratório Nacional de Computação Científica
LNCC - MCTIC**

Relatório de Pesquisa para bolsa PIBIC

**ESTUDO DO MECANISMO DE PROTEÇÃO DE CÉLULAS-TRONCOS CONTRA
APOPTOSE INDUZIDO POR DROGAS**

Aluno: Gabriel Moraes Kleiman

Período da bolsa: 01/2018 - 07/2018

Orientadora: Profa. Dra. Regina Célia Cerqueira de Almeida

Co-orientador: Prof. Dr. Francisco J. P. Lopes

Julho de 2018

I. Introdução

Uma das principais limitações das drogas anticâncer atualmente utilizadas são os mecanismos de proteção das células-tronco cancerígenas contra a ação dessas drogas e sua habilidade de regeneração do tumor. Esses mecanismos são a principal causa da recidiva da doença. Investigamos esse tema usando uma abordagem multidisciplinar, combinando a construção de modelos teóricos com a produção de dados experimentais utilizando a *Drosophila melanogaster* como organismo modelo. O projeto será desenvolvido em colaboração com o Prof. Francisco Lopes, UFRJ – Duque de Caxias, que possui um laboratório de experimentação em *Drosophila*.

O presente estudo visa a produção, solução e ajuste de modelos de redes de sinalização capazes de descrever as alterações no ciclo celular responsáveis pelo estabelecimento do estado de quiescência. Usamos como organismo modelo a *Drosophila melanogaster*. Os estados de quiescência são verificados nas células-tronco germinativas (GSCs – germline stem cells) desse organismo. Buscaremos entender os mecanismos de proteção das GSCs usando uma abordagem multidisciplinar, combinando a produção de dados experimentais e a construção de modelos teóricos e computacionais. Para caracterizar os mecanismos de proteção faremos uma análise morfológica das estruturas dinâmicas chamadas fusomas das GSCs do órgão reprodutor (germaria) do organismo modelo durante o ciclo celular. O fusomo é uma estrutura que faz parte do fuso mitótico das células tronco e apresentam morfologias específicas para cada fase do ciclo celular. Os modelos teóricos visam descrever os mecanismos no nível da regulação gênica. Buscaremos desenvolver modelos de rede de sinalização envolvendo os principais genes envolvidos no ciclo celular, que acreditamos sejam responsáveis pela entrada das células no estado de quiescência.

A identificação dos mecanismos moleculares que levam ao estado de quiescência tem o potencial de contribuir para produção de drogas anticancerígenas além de promover avanços em terapias com células-tronco.

II. Objetivos preestabelecidos

- 1 - Reproduzir a cinética de divisão celular e regeneração do tecido.
- 2 - Identificação dos genes necessários para entrada e saída do estado de proteção.
- 3 - Obtenção de versão paralelizada de um código de AG em uso no laboratório.
- 4 - Modelos teóricos capazes de reproduzir a sinalizar celular e a regulação dos genes que leva aos estados quiescentes.

III. Metas

Fortalecer a colaboração entre os grupos de pesquisa da Profa. Regina de Almeida e do Prof. Francisco Lopes.

Orientar o aluno de iniciação científica (IC) Gabriel Kleiman, DRE 117058447, aluno do curso de Nanotecnologia no Campus UFRJ Duque de Caxias.

Publicar um artigo em revista QUALIS 1, de acordo com a classificação CAPES.

IV. Métodos

- Construção de modelos de redes de sinalização celular
- Solução numérica de modelos e calibração de parâmetros cinéticos a partir de dados experimentais
- Paralelização de códigos de Algoritmos genéticos
- Desafio com drogas (doxorubicina e cisplatina)
- Dissecção de ovários de *Drosophila*
- Imuno-histoquímica
- Microscopia confocal
- Manutenção das moscas desafiadas e observação de suas proles
- Propor modelos teóricos de redes de regulação gênica usando uma estratégia multidisciplinar combinando modelagem teórica (Lopes et al, 2012), análise de imagens (De Sousa, et al 2013) e biologia computacional (Lopes et al, 2008, Holloway, D.M 2011).

V. Resultados e discussão

1. Reproduzir a cinética de divisão celular e regeneração do tecido. - Concluído:

Foram realizados experimentos no Núcleo Multidisciplinar de Pesquisa em Biologia, NUMPEX-BIO, UFRJ Campus Duque de Caxias, no modelo biológico *Drosophila Melanogaster* para avaliar qualitativamente a ação da droga anticâncer cisplatina no tecido dos ovários das fêmeas do modelo. Estes experimentos foram realizados juntamente com a aluna graduanda de biotecnologia deste campus, Ondina Palmares.

2. Identificação dos genes necessários para entrada e saída do estado de proteção. - Não concluído:

Foram identificados muitos genes envolvidos na fase G1 do ciclo celular, da qual é sugerido que seja mais provável a saída da célula do ciclo celular para o estado de quiescência, contudo não houve avanços em identificar entre estes genes os responsáveis diretos por esta saída ou reentrada no ciclo celular. Para avançar nessa etapa será necessário a realização de novos experimentos.

3. Obtenção de versão paralelizada de um código de AG em uso no laboratório. - Objetivo alterado:

Optamos pelo desenvolvimento de uma nova ferramenta matemática para a construção e calibração de modelos de redes de regulação/sinalização, visto que a ferramenta que temos atualmente está escrita em Matlab. Essa nova ferramenta está sendo escrita em uma linguagem aberta, Python 3, estamos implementando diversos recursos visando incluir códigos de algoritmo genético (AG) paralelizados. Desta forma também foi possível aplicar em fase inicial a compatibilidade com modelos biológicos preexistentes em bancos de dados, como Biomodels Database, o que fará da ferramenta apta a ser utilizada por diversos outros grupos de pesquisa futuramente.

Ainda contamos com a colaboração do Professor Adriano Cortês, UFRJ, no

desenvolvimento e otimização desta ferramenta.

Neste primeiro momento obtemos sucesso na leitura do modelo Oregonator, BioModels Database MODEL6622000424, para extração de dados, como número e volume de compartimentos do sistema, espécies envolvidas, reações, concentrações iniciais das espécies e interações entre cada espécie. Com estes dados foi feita automatização da geração das equações referentes a cinética de cada espécie baseada na lei de ação das massas. Optamos por usar este modelo como referência por conta da documentação existente utilizando este modelo com os módulos matemáticos em Python que estamos aplicando.

4. Identificação dos genes envolvidos nos mecanismos de proteção - Objetivo alterado:

Foram estudados os mecanismos de proteção presentes em células-tronco de mamíferos contra moléculas nocivas, como a cisplatina, e foram encontrados mecanismos de reparo de DNA, que podem eliminar pequenos fragmentos de DNA associados a cisplatina. Outros mecanismos são proteínas de bombeamento que conseguem manter uma maior concentração de moléculas no ambiente extracelular. Devido ao redirecionamento adotado no trabalho, decidimos nos concentrar no estudo dos mecanismos em que a molécula NFkB é responsável pela síntese de moléculas envolvidas no câncer de mama humano, pois isto é fundamental para gerar o modelo computacional proposto adiante.

5. Modelos teóricos capazes de reproduzir a sinalizar celular e a regulação dos genes que leva aos estados quiescentes - Objetivo alterado:

As atividades referentes a este objetivo foram desenvolvidas no Núcleo Multidisciplinar de Pesquisa em Computação, NUMPEX-COMP, na UFRJ Campus Duque de Caxias.

A primeira versão deste modelo foi baseada no mapa de sinalização celular presente

no banco de dados *Kyoto Encyclopedia of Genes and Genomes*, *KEGG' PATHWAY map04110*, desta forma selecionamos uma lista dos principais genes envolvidos na fase G1 do ciclo celular e estes foram estudados para gerarmos as equações que descrevem suas interações conformacionais, de síntese, degradação, estímulos e inibições.

Ao finalizar a primeira versão do modelo que descreve a fase G1 do ciclo celular iniciou-se a busca por dados experimentais das concentrações de algumas das moléculas selecionadas para avaliarmos a veracidade do modelo produzido e realizar sua calibração. Contudo, mostrou-se extremamente difícil encontrar estes dados, pois o modelo apresentava um grande número de moléculas, as quais eram analisadas em organismos diferentes em cada fonte encontrada, o que tornava inviável utilizá-las em conjunto por conta das diferenças biológicas entre estes organismos. Foi avaliada a possibilidade de gerar tais dados no Núcleo Multidisciplinar de Pesquisa em Biologia, NUMPEX-BIO, na UFRJ campus Duque de Caxias, utilizando como organismo modelo a *Drosophila Melanogaster*, visto que este modelo já é utilizado pelo Prof. Dr. Francisco J. P. Lopes. Foram feitos orçamentos e consultas com professores da UFRJ acerca do tempo que cada experimento levaria. Este procedimento também se mostrou inviável, tanto por parte financeira quanto por tempo necessário a ser investido.

Frente a inúmeras dificuldades em obter tanto na literatura quanto em bancada os dados experimentais necessários para realizar a calibração do modelo proposto optamos por alterar este objetivo para gerar um modelo capaz de reproduzir a sinalização molecular do NFkB em células de câncer de mama humano, pois em colaboração com o INCA temos a disposição dados experimentais para calibrar este modelo. Como fruto dessa colaboração, recebemos um conjunto de dados experimentais onde foram medidos os níveis de expressão de genes alvos no NKkB como Twist, Slug e SIP.

Uma primeira versão do modelo foi construída e está sendo utilizada a ferramenta matemática *COPASI²* para calibração do modelo com a estimação de parâmetros por algoritmo genético, visto que a ferramenta que estamos produzindo ainda não está operacional.

Atualmente temos dados experimentais ao longo de três pontos no tempo de três das

vinte e quatro moléculas consideradas no sistema. Os primeiros resultados possuem comportamentos aceitáveis tendo em vista o comportamento qualitativo do sistema biológico em relação às espécies que devem ter sua concentração aumentada ou diminuída ao longo do tempo, contudo ainda há dificuldades na questão quantitativa destas espécies. Portanto, ainda estão sendo estudadas as interações entre cada uma dessas espécies em meio biológico para que possamos aproximar de forma quantitativa os resultados obtidos com o esperados. Ainda há a necessidade de maior familiarização com a ferramenta *COPASI* que está sendo utilizada frente ao grande número de recursos que ela apresenta para a pouca documentação disponível.

¹Cell cycle - map04110 - Pathway https://www.genome.jp/dbget-bin/www_bget?map04110

²COPASI: Biochemical System Simulator <http://copasi.org>

VI. Referências Bibliográficas

- Adams, M. D. (2000). The Genome Sequence of *Drosophila melanogaster*. *Science*, 287(5461), 2185–2195. doi:10.1126/science.287.5461.2185
- Cheung-Ong, K., Giaever, G., & Nislow, C. (2013). DNA-Damaging Agents in Cancer Chemotherapy: Serendipity and Chemical Biology. *Chemistry & Biology*, 20(5), 648–659. doi:10.1016/j.chembiol.2013.04.007
- De Sousa, D., Cardoso, M., Bisch, P. M., Lopes, F. JP, and Travençolo, B. (2013). A Segmentation Method for Nuclei Identification from Sagittal Images of *Drosophila melanogaster* Embryos. WSCG, Seite 133-142.
- Donnenberg, V. S., & Donnenberg, A. D. (2005). Multiple Drug Resistance in Cancer Revisited: The Cancer Stem Cell Hypothesis. *The Journal of Clinical Pharmacology*, 45(8), 872–877. doi:10.1177/0091270005276905
- Holloway, D.M., Lopes, Francisco J. P. da Fontoura Costa, Luciano, Travencolo, B. A. N. Golyandina, Nina, Usevich, Konstantin, Spirov, Alexander V. (2011) Gene Expression Noise in Spatial Patterning: hunchback Promoter Structure Affects Noise Amplitude and Distribution in *Drosophila* Segmentation. *PLoS Comput Biol*. 7 (2).
- Lehmann, M., Franco, A., de Souza Prudente Vilar, K., Luíza Reguly, M., & de Andrade, H. H. R. (2003). Doxorubicin and two of its analogues are preferential inducers of homologous recombination compared with mutational events in somatic cells of *Drosophila melanogaster*. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 539(1-2), 167–175. doi:10.1016/s1383-5718(03)00162-1
- Lopes, F. J. P., Vieira, F. M. C., Holloway, D. M., Bisch, P. M., Spirov, A. V. (2008) Spatial Bistability Generates hunchback Expression Sharpness in the *Drosophila* Embryo. *PLoS*

Computational Biology, 4(9). doi: 10.1371/journal.pcbi.1000184.

Lopes, Francisco J.P., Spirov, Alexander V., Bisch, Paulo M. (2012). The role of Bicoid cooperative binding in the patterning of sharp borders in *Drosophila melanogaster*. *Developmental Biology*, 370 (2), 165-172.

NIH Stem Cell Information Home Page. In Stem Cell Information [World Wide Web site]. Bethesda, MD: National Institutes of Health, U.S. Department of Health and Human Services, 2016 [cited November 6, 2016] Available at < [//stemcells.nih.gov/info/basics/7.htm](http://stemcells.nih.gov/info/basics/7.htm)>

Ohlstein, B., Kai, T., Decotto, E., & Spradling, A. (2004). The stem cell niche: theme and variations. *Current Opinion in Cell Biology*, 16(6), 693–699. doi:10.1016/j.ceb.2004.09.003

Orsolin, P. C., Silva-Oliveira, R. G., & Nepomuceno, J. C. (2015). Modulating effect of synthetic statins against damage induced by doxorubicin in somatic cells of *Drosophila melanogaster*. *Food and Chemical Toxicology*, 81, 111–119. doi:10.1016/j.fct.2015.04.004

Podratz, J. L., Staff, N. P., Froemel, D., Wallner, A., Wabnig, F., Bieber, A. J., ... Windebank, A. J. (2011). *Drosophila melanogaster*: A new model to study cisplatin-induced neurotoxicity. *Neurobiology of Disease*, 43(2), 330–337. doi:10.1016/j.nbd.2011.03.022

Pommier, Y., Sordet, O., Antony, S., Hayward, R. L., & Kohn, K. W. (2004). Apoptosis defects and chemotherapy resistance: molecular interaction maps and networks. *Oncogene*, 23(16), 2934–2949. doi:10.1038/sj.onc.1207515

Sawicka, M., Kalinowska, M., Skierski, J., & Lewandowski, W. (2004). A review of selected anti-tumour therapeutic agents and reasons for multidrug resistance occurrence. *Journal of Pharmacy and Pharmacology*, 56(9), 1067–1081. doi:10.1211/0022357044265

Spradling, A., Fuller, M. T., Braun, R. E., & Yoshida, S. (2011). Germline Stem Cells. *Cold Spring Harbor Perspectives in Biology*, 3(11), a002642–a002642. doi:10.1101/cshperspect.a002642

Targeting cancer stem cells: a new therapy to cure cancer patients Yapeng Hu, Liwu Fu *American Journal of Cancer Research (AJCR)*

Yuan, H., & Yamashita, Y. M. (2010). Germline stem cells: stems of the next generation. *Current Opinion in Cell Biology*, 22(6), 730–736. doi:10.1016/j.ceb.2010.08.013

Xie T., Germline stem cell niches (September 30, 2008), StemBook, ed. The Stem Cell Research Community, StemBook, doi/10.3824/stembook.1.23.1, <http://www.stembook.org>. [cited November 6,2016]

Xing, Y., Su, T. T., & Ruohola-Baker, H. (2015). Tie-mediated signal from apoptotic cells protects stem cells in *Drosophila melanogaster*. *Nature Communications*, 6, 7058. doi:10.1038/ncomms8058

0 relatório de atividades deve ser organizado da seguinte forma:

1) Dados gerais

titulo do projeto
nome do bolsista
nome do orientador
tipo de bolsa e período do relatório

2) Objetivos

Incluir sumariamente os objetivos do trabalho realizado.

3) Introdução

A introdução deve conter a informação essencial à compreensão do trabalho.

4) Material e Métodos ou Metodologia

Deve ser sintético mas preciso, contendo explicação dos procedimentos que foram necessários para a execução da pesquisa.

5) Resultados e Discussão

Descrever os resultados obtidos e discuti-los face aos objetivos pretendidos.

6) Conclusões

Sumarizar as principais conclusões obtidas no decurso do trabalho realizado.

7) Referências bibliográficas

Indexação Métrica em Junções Espaciais

Nome do Bolsista: Geovane Noel Alves

Nome do Orientador: Fábio André Machado Porto

Tipo de Bolsa e Período do Relatório: Bolsa de Iniciação Científica - IC
- Julho de 2018

I - Objetivos

Este projeto visa estudar, implementar e avaliar a adoção de índices métricos para auxílio na computação de junções espaciais, implementando nos software Apache Spark e SAVIME.

II - Introdução

Este projeto de iniciação científica se contextualiza na pesquisa por mecanismos eficientes para apoio à busca por padrões geométricos em grandes datasets científicos, como os produzidos por levantamentos astronômicos e aqueles que servem de base para pesquisa sísmica. Nestes datasets, têm-se um grande conjunto de objetos, da ordem de 10^9 , apresentando uma coordenada espacial além de atributos descritivos dos objetos. Por outro lado, buscam-se padrões geométricos especificados por um conjunto de pontos, de mesma natureza dos apresentados nos datasets, que juntos estabelecem uma forma geométrica, a partir de suas posições no espaço. Um exemplo deste tipo de padrão é a Cruz de Einstein (CE) [1]. Este fenômeno, observado por telescópios, reflete a dispersão da luz de um quasar distante quando, em seu caminho para a terra, a luz é desviada por um corpo de grande massa, como uma galáxia. O resultado desta distorção é o aparecimento de 4 sinais luminoso, dispostos no espaço em formato de uma cruz, e que fornecem a ilusão de objetos distintos, quando de fato compartilham a mesma fonte. Buscar formas geométricas que se assemelhem à Cruz de Einstein passou a ser um problema na área de astronomia e em [2] e [3] discutimos técnicas para processar catálogos astronômicos de forma eficiente, em busca de variações da CE.

Em [3], discutimos uma técnica em que, inicialmente, para cada objeto do catálogo, procura-se os $k-1$ objetos candidatos a comporem uma forma geométrica coincidentes com a forma sendo procurada, para uma forma com k pontos. Uma vez encontrados, os objetos candidatos são combinados aplicando-se um algoritmo de junção, que valida a distância par-a-par entre os objetos candidatos à solução, e as distâncias entre os objetos dados como o padrão de busca. Nossa implementação adota um algoritmo de junção de loops alinhados, o que é reconhecidamente ineficiente. Por outro lado, é conhecida a proposta de indexação através de estruturas métricas [4]. Índices métricos organizam um conjunto de dados a partir das distâncias a um ponto do conjunto c . Usando-se da desigualdade triangular, podem-se obter todos os objetos a uma distância d_i de um segundo objeto o , usando-se as distâncias conhecidas ao objeto c . Desta forma, deseja-se avaliar a utilização de tais estruturas de indexação para melhorar a eficiência do processo de junção espacial na solução do problema de busca de formas espaciais em grandes bases de dados. A proposta deverá ser integrada à implementação, Constellation Queries, desenvolvida em Apache Spark, no laboratório DEXL.

III - Material e Métodos ou Metodologia

Neste período de iniciação científica foram realizadas as seguintes tarefas:

- a) aprendizado da linguagem Python, sendo testado no editor de texto em conjunto com o terminal do computador.
- b) aprendizado sobre a Slim-Tree, sendo estudada pelo artigo [5];
- c) apresentação do desenvolvimento do trabalho para o grupo DEXL Lab.

IV - Resultados e Discussão

O estudo da Slim-Tree foi feita com base nos artigos [4] e [5]. Com estes foi visto como esta árvore é construída, como se comporta e seus algoritmos de análise e otimização. O estudo do Python foi dado por um site de cursos online, foi estudado todo o ciclo básico do Python.

A Slim-Tree é uma árvore multivias balanceada e dinâmica desenvolvida para indexar objetos contidos em um espaço métrico. Utiliza um método de acesso métrico (MAM) para dividir o espaço em regiões e sub-regiões não disjuntas com o intuito de maximizar a velocidade de criação e de pesquisa.

Utilizando um algoritmo de análise para determinar quão boa é a árvore, Fat-Factor, e um de otimização, Slim-Down. Também utiliza um algoritmo de visualização, FastMap, para facilitar o entendimento de como a árvore é construída e utilizada.

V - Conclusão

Neste período foi estudado alguns conceitos básicos sobre Python. Aprendendo também algumas curiosidades, sendo um exemplo como funciona a indentação em Python, sendo que os algoritmos só funcionam se forem corretamente indentados e a falta de ponto e vírgula(;) ao final de cada linha de código em Python, cujo me deixou bem confuso no início.

Outro assunto que foi estudado foi como funciona a Slim-Tree, sendo uma árvore multivias, balanceada que seu foco é em aumentar o desempenho nas pesquisas, reduzir o número de acessos ao disco, permitir inserção de dados a qualquer momento da vida útil da árvore e seus algoritmos de análise(Fat-Factor) e otimização(Slim-Down). Fat-Factor analisa a árvore e calcula o grau de sobreposição entre os nós da árvore e gera um valor inteiro positivo para representar este grau, sendo 0 como nenhum. Quando o grau de sobreposição entre os nós da árvore está muito alto utiliza-se o Slim-Down, que consiste em pegar os LeafNode vinculados ao mesmo IndexNode e rearranja-los de forma a reduzir o raio de cobertura e o grau de sobreposição entre estes nós.

Estudando neste período foi percebido que a linguagem Python é uma linguagem de programação simples e poderosa, que foge um pouco dos padrões de outras linguagens, como C++ e Java, e obriga o programador a fazer a indentação corretamente. Já a Slim-Tree se utiliza de sua velocidade superior em relação a outras árvores, tanto na criação quanto na pesquisa por respostas pontuais, sendo ainda mais versátil quando utilizada depois de ser otimizada pelo seu algoritmo.

VI - Referências Bibliográficas

[1] Huchra, J., Gorenstein, M., Kent, S., et al., A new and unusual gravitational lens, *Astronomical Journal* (ISSN 0004-6256), vol. 90, May 1985, p. 691-696.

[2] Amir Khatibi, Fábio Porto, João Guilherme Rittmeyer, Eduardo S. Ogasawara, Patrick Valduriez, Dennis, Shasha: Pre-processing and Indexing Techniques for Constellation Queries in Big Data. *DaWaK 2017*: 164-172

[3] Fábio Porto, Amir Khatibi, João R. Nobre, Eduardo S. Ogasawara, Patrick Valduriez, Dennis E. Shasha: Constellation Queries over Big Data. *ArXiv:/1703.02638* (2017)

[4] Traina, C, Traina, A., Faloutsos, C., Fast Indexing and Visualization of Metric Data sets using slim-trees, *IEEE Transactions in Knowledge and Engineering*, V(14), I(2)

[5] Traina, Chino, Visualizando a organização e o comportamento de estruturas métricas: Aplicações em consultas por similaridade, USP - São Carlos Março/2004

RELATÓRIO DE ATIVIDADES DA XIII JORNADA DE INICIAÇÃO CIENTÍFICA E TECNOLÓGICA

BOLSA PIBIC/LNCC

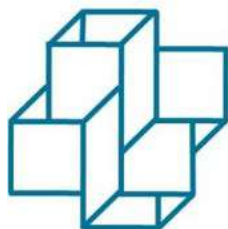
Localização de Robôs Móveis via Filtro de Kalman

Bolsista: Hiago Riba Guedes

Orientador: Marcelo Dutra Fragoso

Período: 01/08/2017 a 31/07/2018

Projeto: 800333/2016-0



Petrópolis
19 de Julho de 2018

1 Objetivos

Este trabalho tem como objetivos a *confeção* de um robô móvel e a implementação de ferramentas adequadas para a *localização* desse robô. O problema de localização de robôs é fundamental na construção de robôs autônomos. Nesse contexto, o Filtro de Kalman tem um papel importante na obtenção de estimativas ótimas, no sentido da média quadrática, dos dados ruidosos.

As principais etapas contempladas pelo Projeto pode ser subdividida da seguinte maneira:

- Leitura da literatura sobre localização de robôs móveis,
- Estudo teórico sobre os filtros de Kalman
- Realização da montagem do robô
- Aplicação do Filtro de Kalman no problema de localização.
- Aplicação dos dados a um controle PID

2 Introdução

Devido, em parte, a importância da utilização de robôs em várias aplicações, existe atualmente um esforço considerável de pesquisa nessa área, que pode ser comprovado pelo número de artigos que têm sido publicados recentemente sobre o assunto. Robôs tem tido um papel importante no desenvolvimento da indústria moderna. Além disso, robôs autônomos têm sido utilizados, por exemplo, para exploração de ambientes que são de difícil acesso ao ser humano; missões espaciais (ex., "the lunar rover initiative"); exploração em águas profundas; setor de serviços, etc. Eles tem sido também aplicados nas forças armadas e em inúmeros projetos acadêmicos onde pode ser destacado o futebol de robôs.

Esse trabalho tem como objetivo, além da confecção do robô, o estudo de uma das principais etapas do problema de localização, que é o tratamento de dados do robô, uma vez que os dados de entrada tendem a ser muito ruidosos por conta da própria arquitetura eletrônica presente do robô e do próprio ambiente em que ele se encontra, onde existe cada vez mais potenciais causadores de interferência para o sistema. Para isso precisaremos aliar os sensores proprioceptivos e heteroceptivos para conseguirmos a execução ótima de determinada função que dependa de sua localização. Vale salientar que, a necessidade de uma localização precisa faz com que sistema como o GPS, por apresentar baixa precisão, tornar-se inviável.

A priori o robô terá inteligência simples, sendo de competência própria apenas o desvio de obstáculos e com isso ele deverá nos fornecer em que ponto da localidade ele se encontra em relação ao seu ponto de origem e, via Bluetooth, nos oferecer suas coordenadas locais.

3 Materiais e Métodos

Seguindo os passos estipulados pelo objetivo temos:

3.1 Montagem do Robô

Para a confecção do hardware do robô foi utilizado um microcontrolador com arquitetura AVR, onde a eles confere o objetivo de adquirir os dados dos sensores presentes, tais como sensores de velocidade, giroscópio e ultrassom e fazer a sua devida filtragem com o objetivo final de ter conhecimento da sua localização ótima. 4 Baterias do tipo 18650 foram utilizados a fim de termos uma tensão de alimentação próxima aos 15 V, suficientes para dar uma autonomia razoável para o mesmo, contamos também com

uma placa responsável para a proteção e carregamento adequado das baterias. Um esquema relativo ao hardware do robô, assim como uma foto do robô montado, aparecem nas figuras 1 e 2, respectivamente.

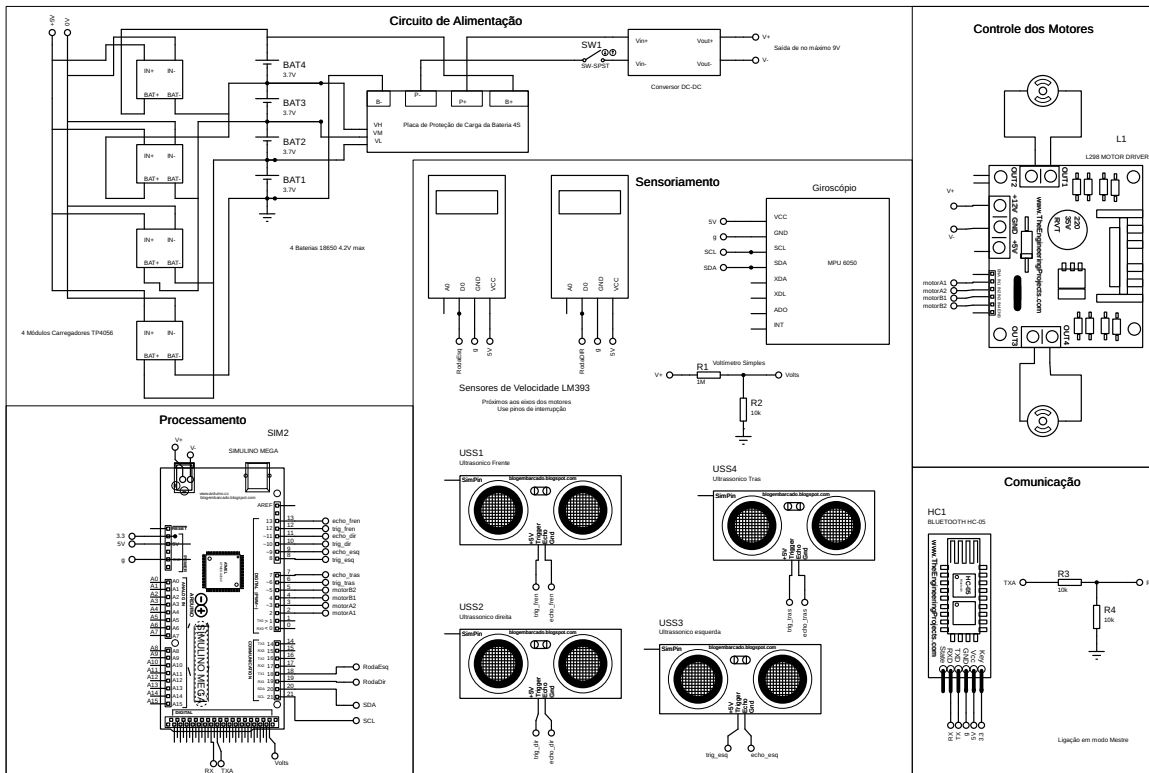


Figura 1: Esquemático do hardware do Robô

3.2 O Filtro de Kalman

Conhecido por ser um dos muitos responsáveis por levar o homem a Lua no projeto Apollo, o Filtro de Kalman tem sido utilizados na implementação de sistemas de navegação de submarinos; guiamento e navegação de mísseis, etc. Além disso, tem sido utilizados em problemas oriundos da meteorologia, computação gráfica e um grande número de outros problemas onde se faz necessário obter estimativas ótimas de sinais corrompidos por ruídos. Por ser um algoritmo recursivo e por atuar no erro médio quadrado esse filtro se mostra bastante prático para resolver o problema de ruído que ocorre no sensoriamento.

Na sua forma mais geral, o Filtro de Kalman tem como objetivo calcular estimativas ótimas, no sentido de minimizar o erro médio quadrática das estimativas, da variável de estado do seguinte sistemas dinâmico à tempo-discreto:

$$\begin{aligned}
 x_{k+1} &= A_k x_k + B_k u_k + C_k w_k \\
 z_k &= H_k x_k + G_k v_k,
 \end{aligned}$$

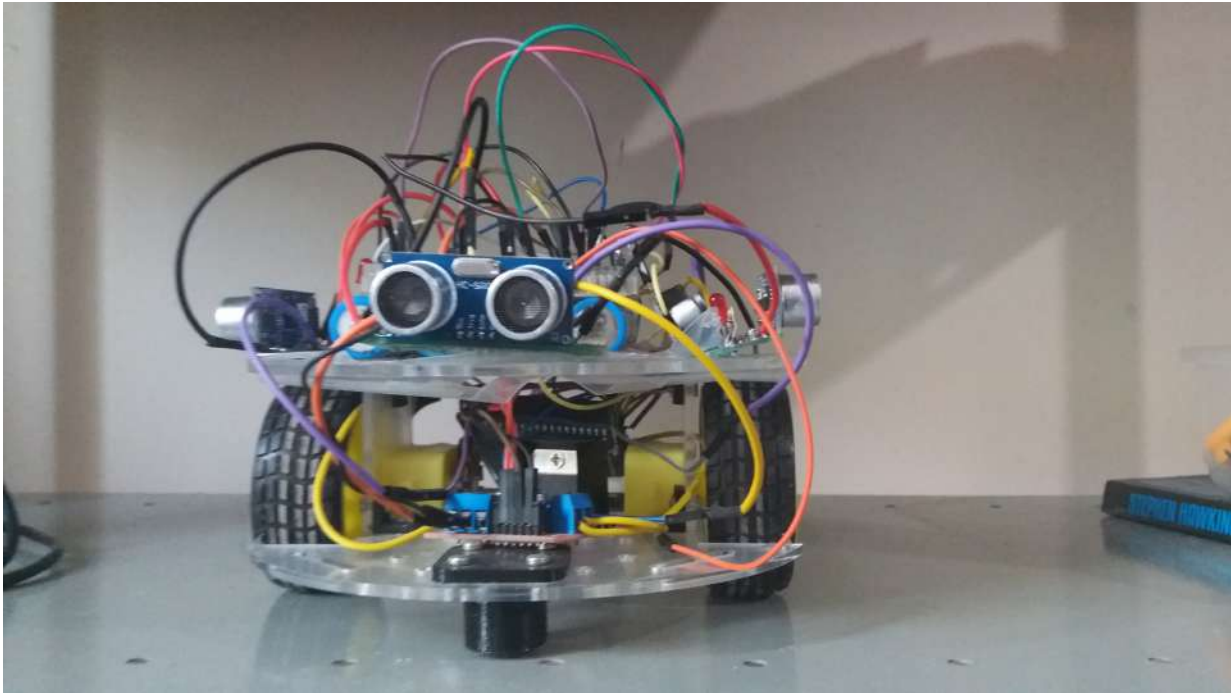


Figura 2: Foto do Robô montado

onde assumimos que x_0 , $\{w_k\}$ e $\{v_k\}$ são mutuamente independentes, com $\{w_k\}$ e $\{v_k\}$ Gaussianos com média zero e covariância unitária. O filtro de Kalman é dado então por:

Equação de predição de estado a priori

$$\hat{x}_k^- = A\hat{x}_{k-1}^+ + Bu_k$$

$$P_k^- = AP_{k-1}^+A^T + C_kQC_k^T$$

Agora mostrando as equações a posteriori

$$\hat{x}_k^+ = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-)$$

$$P_k^+ = P_k^- - K_kH_kP_k^-$$

$$K_k = P_k^-H_k^T(H_kP_k^-H_k^T + G_kR_kG_k^T)^{-1}$$

Onde:

A_k Matriz do sistema

B_k Matriz de Controle

P_k Matriz de Covariância do Erro ($[x_k - \hat{x}_k][x_k - \hat{x}_k]^T$)

z_k Vetor de saída dos sensores, vetor que contém o ruído que desejamos retirar

Q_k Matriz de Covariância do Ruído do Ambiente

R_k Matriz de Covariância do Ruído de saída

$\hat{x}_k = E[x_k | F^y]$

$F_k^y = \sigma[y_1, y_2, \dots, y_k]$

Onde entraremos com nossos parâmetros pertinentes na entrada, que seriam as coordenadas dadas pelos nossos sensores (que no caso são as coordenadas x, y e a orientação θ). Vale salientar que a forma como o filtro de Kalman foi acima apresentada, esta relacionada com o fato da sua melhor utilização em sistemas embarcados, além do fato de conseguirmos monitorar seus parâmetros. No entanto, fazendo-se as devidas substituições resgatamos o filtro de Kalman em sua forma canônica.

3.3 Os Modelos Utilizados

Foram utilizados dois modelos para estudar o filtro de Kalman no robô móvel, o *modelo linear* que é normalmente denominado na literatura de Modelo Cinemático, que é assim denominado pois somente nos interessa como o robô irá se comportar geometricamente após se movimentar por um certo período de tempo e o *modelo não-linear* que também é denominado de modelo Dinâmico, que é assim denominado por levar em consideração qualquer força vinculada a sua movimentação.

3.3.1 Determinação das Matrizes no sistema Cinemático

Após a análise de um desenho representativo do robô em um momento k e um momento $k+1$ e combiná-las com equações básicas de movimento de um robô conseguimos estipular o seguinte modelo :

$$A_k = \begin{bmatrix} x_k & 0 & \Delta s \cos(\theta + \frac{\Delta\theta}{2}) \\ 0 & y_k & \Delta s \sin(\theta + \frac{\Delta\theta}{2}) \\ 0 & 0 & \theta_0 + \frac{\Delta\theta}{2} \end{bmatrix} = \begin{bmatrix} x_k & 0 & \frac{\Delta s_r + \Delta s_l}{2} \cos(\theta + \frac{\Delta s_r - \Delta s_l}{2b}) \\ 0 & y_k & \frac{\Delta s_r + \Delta s_l}{2} \sin(\theta + \frac{\Delta s_r - \Delta s_l}{2b}) \\ 0 & 0 & \theta_0 + \left(\frac{\Delta s_r - \Delta s_l}{2b}\right) \end{bmatrix}$$

Como o robô terá uma inteligência um tanto rudimentar como não bater em nenhuma parede, a matriz de modelo das medidas será a matriz identidade dos estados.

$$H_k = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

3.3.2 Determinação da Matriz no sistema Dinâmico

Por se tratar de uma abordagem não muito comum para robôs diferenciais, a abordagem não linear acaba por ser não muito utilizada na literatura. Muito pela dificuldade em se adquirir os dados necessários para o modelo.

$$\frac{\partial f}{\partial x} = \begin{bmatrix} 1 & 0 & -\frac{1}{2}T(\gamma_l \omega_l + \gamma_r \omega_r) \sin(\theta) \\ 0 & 1 & \frac{1}{2}T(\gamma_l \omega_l + \gamma_r \omega_r) \cos(\theta) \\ 0 & 0 & 1 \end{bmatrix}$$

4 Resultados

Ao longo dos 12 meses, as seguintes etapas do Projeto foram realizadas: (1) Leitura da literatura sobre localização de robôs móveis; (2) Seminários sobre a Teoria de Filtros de Kalman; (3) Realização da montagem do robô; (4) Estudo dos modelos para o robô e identificação da variáveis ruidosas; (5) Confecção preliminar do código relativo à aplicação do filtro de Kalman.

Para averiguar a eficácia dos modelos encontrados serão realizados dois testes, sendo o primeiro um teste que é utilizado pela DARPA (Defense Advanced Research Projects Agency) para premiar o carro autônomo mais preciso que é o de fazer com que o carro dê uma volta completa em um circuito de distância conhecida e pare exatamente no mesmo local em que começou. O carro que apresentar menor erro real é considerado o vencedor. Faremos aqui uma versão mais simples porém com a mesma ideia, esticamos uma fita isolante no chão e medimos o comprimento total dessa fita. Faremos com que o robô

ande essa distância sem e com o Filtro de Kalman em seu código. Após esse teste faremos com que o robô dê voltas em uma trajetória circular , com um raio de giro conhecido. Com isso extraíremos valores com e sem o filtro.

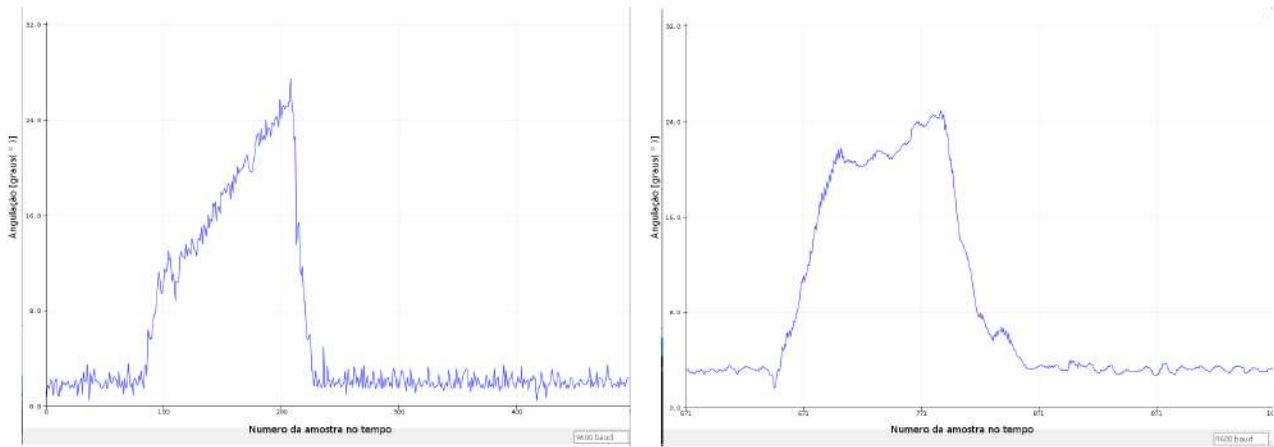


Figura 3: À esquerda temos o valor de angulação fornecido pelo giroscópio sem o filtro. À direita temos o resultado aplicando o filtro, onde observa-se uma atenuação do ruído.

Até o momento, conseguimos aplicar um filtro específico para o giroscópio, e nele pudemos notar uma melhora significativa em suas medidas. Conseguimos também fazer um código usando o seu modelo linear . Fazendo com que o robô dê uma volta em torno de um raio teoricamente constante.

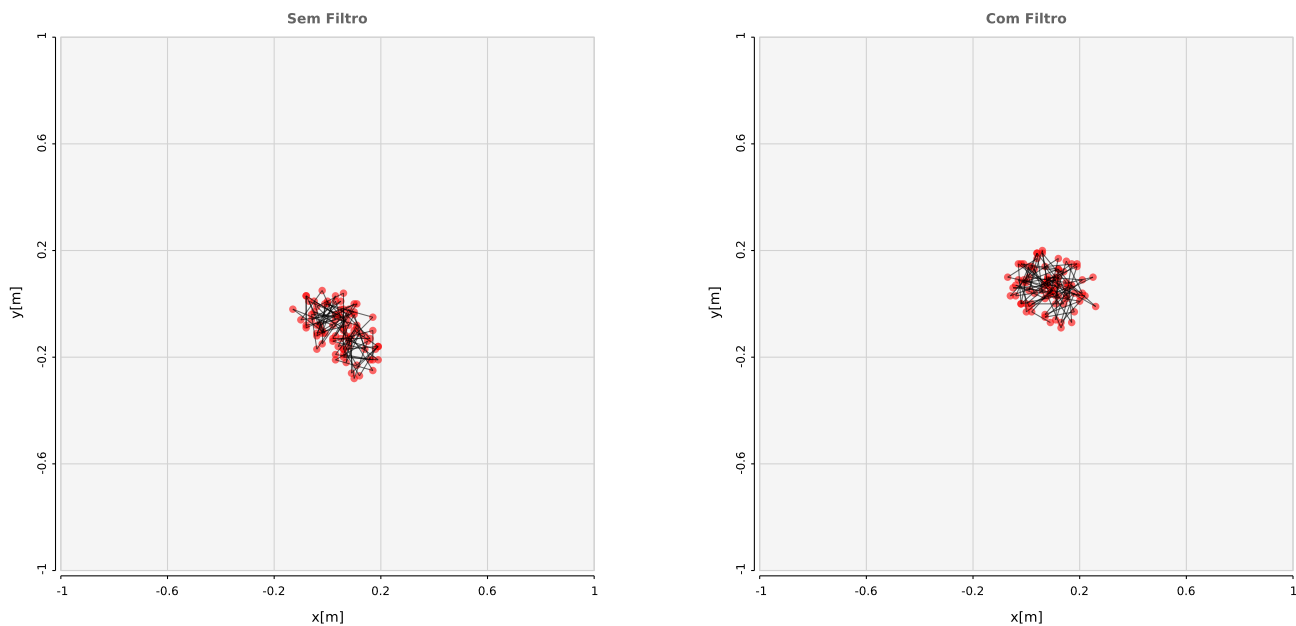


Figura 4: À esquerda temos a localização do robô sem o filtro e a direita temos a aplicação do filtro. Os resultados ainda se apresentam de forma bastante aleatória porém apresenta uma conformidade mais realista com a movimentação do robô ,a imprecisão se deve bastante em conta dos sensores utilizados (Um disco encoder de 10 furos).

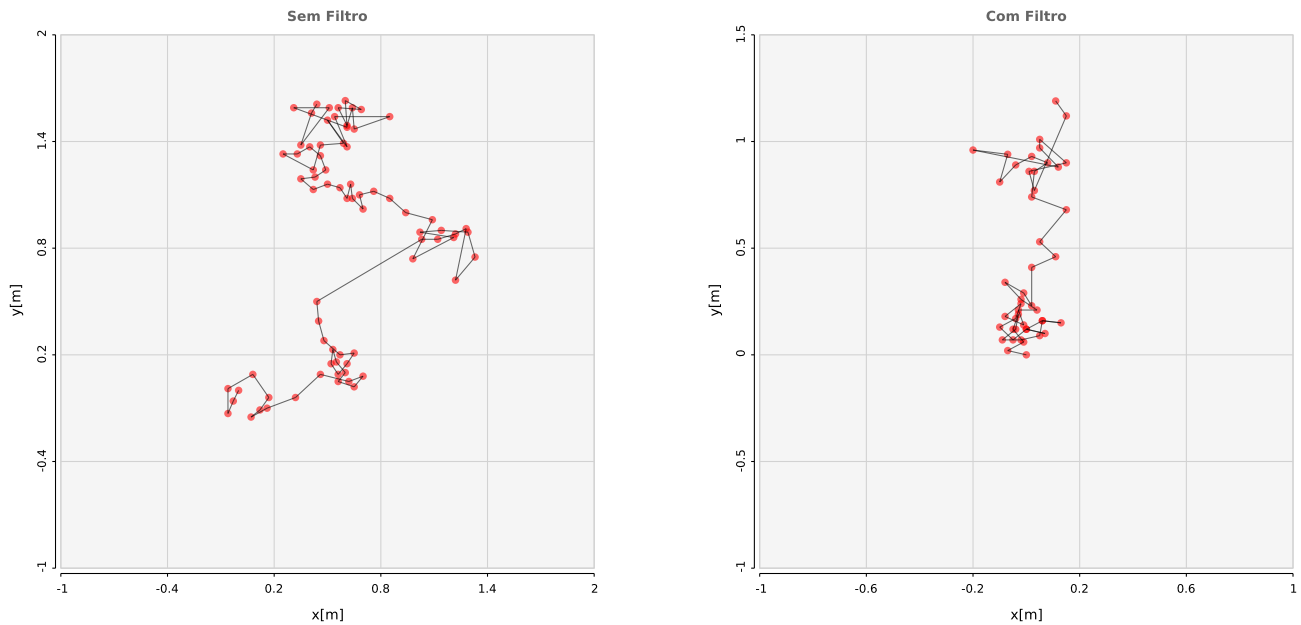


Figura 5: À esquerda temos o movimento do robô em linha reta sem o filtro e à direita temos uma movimentação em linha reta do robô com a aplicação do filtro.

5 Conclusões e Próximos Trabalhos

Ao longo do período em questão foram realizadas diversas etapas do Projeto. A primeira etapa envolveu o estudo da literatura sobre localização de robôs móveis e seminários sobre filtros de Kalman. A segunda etapa tratou da montagem do robô, estudo do modelo e identificação das variáveis de estado ruidosas. A terceira etapa foi focada na confecção do código. A etapa final da aplicação está em estágio de finalização. Dentre os diversos problemas encontrados, um dos mais significativos é referente à determinação das estatísticas do ruído.

Além da necessidade de finalização da aplicação, seria interessante dedicar mais tempo a questões importante na aplicação do filtro de Kalman. Por exemplo, um melhor entendimento do que tem sido chamado de "sensor fusion" poderá ser útil na aplicação. Além disso, dada a dificuldade em lidar com a questão das estatísticas do ruído, seria interessante tentar utilizar a versão do filtro de Kalman robusto, isto é, onde considera-se como ruído qualquer sequência de variáveis aleatórias pertencente ao l_2 . Finalmente, seria interessante verificar o tipo de ganho ao se utilizar um filtro de partícula.

Referências

- [1] Robot Localization and Kalman Filters *On finding your position in a noisy world*, Rudy Negenborn, UTRECHT UNIVERSITY Link: http://www.negenborn.net/kal_loc/thesis.pdf
- [2] Stochastic Modelling and Control, M. H. A. DAVIS and R. B. VINTER, ISBN-13: 978-94-010-8640-0
- [3] Introdução à Robótica, Maja J. Mataric, ISBN: 978-85-393-0490-5
- [4] Introduction to Kalman Filters and SLAM, R. Siegwart and I. Nourbakhsh P. Rybski G. Welch and G. Bishop Link: <http://web.eecs.utk.edu/~leparkar/Courses/CS594-fall108/Lectures/Nov-20-Localization-Mapping-III.pdf>

Relatório de atividades:

Métodos de Discretização para Sistemas Físicos Dispersivos e/ou Dissipativos

Aluno: Igor P. dos Santos Pereira

Orientadora: Sandra M. Cardoso Malta

Co-orientadora: Cristiane Oliveira de Faria

20 de Julho de 2018

1 Objetivo

Propor métodos explícitos de Diferenças Finitas para a equação de Korteweg–de Vries linearizada.

2 Introdução

Em 1834, o engenheiro britânico John Scott Russel observou no canal de Eddinburg-Glasgow uma elevação de água capaz de viajar por longa distância em um canal raso, com velocidade constante sem se dissipar, fenômeno o qual foi batizado de “onda solitária” [1, 2]. Posteriormente, os holandeses Korteweg e de Vries deduziram, a partir das equações de Navier-Stokes [6], a EDP que modelava o fenômeno observado por Russel. A versão adimensionalizada desta equação passou a ser conhecida como Equação de Korteweg–de Vries (KdV), a qual pode ser escrita da seguinte forma:

$$u_t + uu_x + u_{xxx} = 0. \quad (1)$$

Esta equação inclui tanto efeitos dispersivos, oriundos de sua derivada de terceira ordem u_{xxx} , quanto efeitos não-lineares de uu_x . Enquanto o termo dispersivo tende a atenuar a oscilação da solução, o termo não linear proporciona a formação de choque, isto é, a perda de continuidade da solução devido a oscilações abruptas. Sendo assim, o equilíbrio dessas duas “forças” opostas tende a gerar soluções estáveis, sem descontinuidades: as já citadas, “ondas solitárias” ou *sólitons* [2].

Após uma colisão, os dois sólitons permanecem com mesma amplitude e velocidade, apenas sofrem uma defasagem de suas posições em relação às que teriam, no mesmo instante do tempo, caso estivessem isolados. Justamente por preservar seu formato após a colisão, que as ondas solitárias também são conhecidas pelo nome de *sólitons*, que remete ao comportamento de partículas [2].

De certa forma, a KdV pode ser observada como um caso particular de uma classe de equações mais geral. A Equação de Kawahara é uma EDP de quinta ordem, não-linear, que modela uma série de fenômenos dispersivos, ondas de plasma, ondas capilares, entre outras [10]:

$$u_t + uu_x + \alpha u_{xxx} + \beta u_{xxxx} = 0. \quad (2)$$

Decifrar cada termo da equação (2) é compreender melhor os fenômenos físicos por trás desses modelo. Nesse sentido, como um primeiro passo, a investigação da derivada cúbica u_{xxx} nesse modelo. Desprezando-se o termo não-linear uu_x de (1) e acrescentando um termo fonte, obtém-se aquela que neste trabalho chamaremos de Equação de Korteweg–de Vries Linearizada [1]. A saber,

$$u_t + u_{xxx} = f. \quad (3)$$

Esta equação modela a propagação de ondas rasas de água e longas ondas de água com pequenas amplitudes [1]. É de se esperar que a falta da não-linearidade em (4) faça com que suas soluções sejam somente influenciadas pelo caráter dispersivo da derivada terceira u_{xxx} . Esse fato pode ser verificado, por exemplo, utilizando métodos numéricos para determinação de uma solução aproximada para a KdV linearizada. Em particular, o Método de Diferenças Finitas foi o escolhido neste trabalho para realizar tal verificação.

O problema modelo adotado, baseado em [9], pode ser escrito como:

$$\begin{cases} u_t + u_{xxx} = f , & \forall (x, t) \in \Omega = (0, L) \times (0, T), \\ u(0, t) = u(L, t) = 0 , & \forall t \in [0, T], \\ u_x(L, t) = 0 , & \forall t \in [0, T], \\ u(x, 0) = g(x) , & \forall x \in [0, L]. \end{cases} \quad (4)$$

onde, $f \in C^0(\Omega)$ é o termo fonte, $g \in C^3(0, L)$ é a condição inicial do problema, $T > 0$ e $L > 0$.

3 Metodologia

Os métodos de diferenças finitas utilizados para estudar a KdV linearizada (4) são baseados em métodos clássicos para a equação da advecção. A equação de advecção é a equação diferencial parcial mais simples para modelar sistemas hiperbólicos, a qual pode ser escrita da seguinte forma

$$u_t + au_x = 0 , \quad a > 0 .$$

Nesse contexto, pode-se destacar três métodos explícitos de 1-passo: método de Euler, método de Lax-Friedrichs e método de Lax-Wendroff [4, 7, 8]. Alguns esquemas de diferenças finitas propostos nesse trabalho se inspiram nesses três métodos. Por questões de conveniência, serão referidos pelos mesmos nomes. Além desses três métodos, foi proposto um quarto método que utiliza uma discretização do termo u_{xxx} como sugerido em [5].

A dedução de cada um dos esquemas de diferenças finitas será feita a seguir. Em todos os casos, a malha é construída de forma uniforme.

3.1 Método de Euler explícito – central (EE-C)

O método de Euler consiste em utilizar um esquema progressivo para discretizar a derivada temporal u_t , como mostrado a seguir

$$\frac{\partial u}{\partial t} \Big|_i^n = \frac{u_i^{n+1} - u_i^n}{\Delta t} .$$

O termo da derivada terceira u_{xxx} , discretizado com o seguinte esquema central:

$$\frac{\partial^3 u}{\partial x^3} \Big|_i^n = \frac{u_{i+2}^n - 2u_{i+1}^n + 2u_{i-1}^n - u_{i-2}^n}{2\Delta x^3} .$$

Assim, propõe-se um esquema de diferenças finitas para a equação (4), avaliando o termo fonte em (x_i, t_n) , isto é $f(x_i, t_n) = f_i^n$. É possível explicitar o termo do passo $n + 1$ em função de todos os outros, do passo n , uma vez que o método é explícito:

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{2\Delta x^3} \left(u_{i+2}^n - 2u_{i+1}^n + 2u_{i-1}^n - u_{i-2}^n \right) + \Delta t f_i^n \quad (5)$$

3.2 Método de Euler explícito – Holden (EE-H)

Nesse método, é utilizado o mesmo tipo de discretização do termo u_t : um esquema de derivada progressiva. A diferença deste para o método anterior está na discretização da derivada de terceira ordem. Em [5] os autores definem os seguintes operadores de diferenças finitas:

$$D_{\pm}p(x) = \pm \frac{1}{\Delta x} \left(p(x \pm \Delta x) - p(x) \right).$$

Com isso, pode-se compor esses operadores a fim de se obter um operador de discretização para a derivada terceira:

$$D_+^2 D_- u_i^n = \frac{u_{i+2}^n - 3 u_{i+1}^n + 3 u_i^n - u_{i-1}^n}{\Delta x^3}.$$

O que nos leva a um novo esquema de diferenças finitas para a KdV linearizada:

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{\Delta x^3} \left(u_{i+2}^n - 3 u_{i+1}^n + 3 u_i^n - u_{i-1}^n \right) + \Delta t f_i^n \quad (6)$$

3.3 Método de Lax-Friedrichs (LF)

O método de Lax-Friedrichs parte do método de Euler explícito e está fundamentado na ideia de substituir o termo u_i^n de (5) pela média aritmética da solução numérica nos pontos adjacentes (x_{i+1}, t_n) e (x_{i-1}, t_n) :

$$u_i^{n+1} = \frac{1}{2} \left(u_{i+1}^n + u_{i-1}^n \right) - \frac{\Delta t}{2\Delta x^3} \left(u_{i+2}^n - 2 u_{i+1}^n + 2 u_{i-1}^n - u_{i-2}^n \right) + \Delta t f_i^n \quad (7)$$

3.4 Método de Lax-Wendroff (LW)

O método de Lax-Wendroff é deduzido obtendo uma equação de ordem mais alta a partir da equação original:

$$\begin{aligned} \frac{\partial u}{\partial t} + \frac{\partial^3 u}{\partial x^3} &= f(x, t) \\ \frac{\partial}{\partial t} \left(\frac{\partial u}{\partial t} \right) &= \frac{\partial}{\partial t} \left(f(x, t) - \frac{\partial^3 u}{\partial x^3} \right) = \frac{\partial f}{\partial t}(x, t) - \frac{\partial^3}{\partial x^3} \left(\frac{\partial u}{\partial t} \right) = \frac{\partial f}{\partial t}(x, t) - \frac{\partial^3}{\partial x^3} \left(f(x, t) - \frac{\partial^3 u}{\partial x^3} \right) \\ \frac{\partial^2 u}{\partial t^2} &= \frac{\partial f}{\partial t}(x, t) - \frac{\partial^3 f}{\partial x^3}(x, t) + \frac{\partial^6 u}{\partial x^6} \end{aligned}$$

A ideia por trás dessa manipulação é gerar outra relação entre as derivadas de x e t – além da equação original – e usá-las numa expansão em série de Taylor para a variável t , truncada em segunda ordem:

$$\begin{aligned} u(x_i, t_n + \Delta t) &= u(x_i, t_n) + \Delta t \frac{\partial u}{\partial t}(x_i, t_n) + \frac{\Delta t^2}{2!} \frac{\partial^2 u}{\partial t^2}(x_i, t_n) + O(\Delta t^3) \\ u(x_i, t_n + \Delta t) &= u(x_i, t_n) + \Delta t \left(f(x_i, t_n) - \frac{\partial^3 u}{\partial x^3}(x_i, t_n) \right) + \frac{\Delta t^2}{2!} \left(\frac{\partial f}{\partial t}(x_i, t_n) - \frac{\partial^3 f}{\partial x^3}(x_i, t_n) + \frac{\partial^6 u}{\partial x^6}(x_i, t_n) \right) + O(\Delta t^3) \\ u_i^{n+1} &= u_i^n - \Delta t \frac{\partial^3 u}{\partial x^3} \Big|_i^n + \frac{\Delta t^2}{2} \frac{\partial^6 u}{\partial t^6} \Big|_i^n + \Delta t f_i^n + \frac{\Delta t^2}{2} \left(\frac{\partial f}{\partial t} \Big|_i^n - \frac{\partial^3 f}{\partial x^3} \Big|_i^n \right) \end{aligned}$$

Discretizando as derivadas de u com esquemas centrais:

$$\begin{aligned} u_i^{n+1} &= u_i^n - \frac{\Delta t}{2\Delta x^3} \left(u_{i+2}^n - 2 u_{i+1}^n + 2 u_{i-1}^n - u_{i-2}^n \right) \\ &\quad + \frac{\Delta t^2}{2\Delta x^6} \left(u_{i+3}^n - 6 u_{i+2}^n + 15 u_{i+1}^n - 20 u_i^n + 15 u_{i-1}^n - 6 u_{i-2}^n + u_{i-3}^n \right) \\ &\quad + \Delta t f_i^n + \frac{\Delta t^2}{2} \left(\frac{\partial f}{\partial t} \Big|_i^n - \frac{\partial^3 f}{\partial x^3} \Big|_i^n \right) \end{aligned} \quad (8)$$

3.5 Termo fonte, condição inicial e tratamento das condições de contorno

Em todas simulações foram utilizadas as seguintes escolhas para a condição inicial e o termo fonte:

$$g(x) = \frac{x(x-L)^2}{L^2}, \quad f(x, t) = \frac{6 - x(x-L)^2}{L^2} e^{-t} \quad \Rightarrow \quad \frac{\partial f}{\partial t}(x, t) - \frac{\partial^3 f}{\partial x^3}(x, t) = \frac{x(x-L)^2}{L^2} e^{-t}$$

Uma vez que $x_i = i\Delta x$ e $t_n = n\Delta t$, é possível discretizar a condição inicial e o termo fonte da seguinte forma:

$$\begin{aligned} u_i^0 &= g(x_i), & \forall i \\ f_i^n &= f(x_i, t_n), & \forall i, n \\ \frac{\partial f}{\partial t} \Big|_i^n - \frac{\partial^3 f}{\partial x^3} \Big|_i^n &= \frac{\partial f}{\partial t}(x_i, t_n) - \frac{\partial^3 f}{\partial x^3}(x_i, t_n), & \forall i, n \end{aligned}$$

O mesmo pode se feito para as condições de Dirichlet em $x = 0$ e $x = L$:

$$u_0^n = u_I^n = 0, \quad \forall n, \quad x_I = L.$$

E também para a condição de contorno de Neumann:

$$\frac{\partial u}{\partial x} \Big|_I^n = \frac{-3u_I^n + 4u_{I-1}^n - u_{I-2}^n}{2\Delta x} = 0 \quad \therefore \quad u_{I-1}^n = \frac{1}{4}u_{I-2}^n, \quad \forall n$$

Nas vizinhanças de $x = 0$, mais precisamente em $i = 1$, nenhum dos esquemas propostos possui validade. Sendo assim, é necessário o uso de um esquema auxiliar nessa região para que a solução numérica seja obtida:

$$u_1^{n+1} = u_1^n - \frac{\Delta t}{2\Delta x^3} \left(-u_4^n + 6u_3^n - 12u_2^n + 10u_1^n - 3u_0^n \right) + \Delta t f_1^n \quad \{EEC/EEH/LF/LW\}$$

Esse esquema pode ser usado juntamente com todos os métodos propostos. Entretanto, no caso do Método de Lax-Wendroff, o esquema auxiliar precisa ser utilizado também em $i = 2$: o esquema de Lax-Wendroff possui o maior estêncil entre todos os métodos estudados. Em vista disso, a necessidade do uso do esquema auxiliar em mais de um ponto.

$$u_2^{n+1} = u_2^n - \frac{\Delta t}{2\Delta x^3} \left(-u_5^n + 6u_4^n - 12u_3^n + 10u_2^n - 3u_1^n \right) + \Delta t f_2^n \quad \{LW\}$$

Por outro lado, nas vizinhanças de $x = L$, apenas Lax-Wendroff não pode ser utilizado em $i = I - 2$, pelo mesmo motivo explicado no caso de $i = 2$. Portanto, é necessário o uso de outro esquema auxiliar. Nesse caso, aplica-se o Método de Euler explícito – central para esse ponto:

$$u_{I-2}^{n+1} = u_{I-2}^n - \frac{\Delta t}{2\Delta x^3} \left(u_I^n - 2u_{I-1}^n + 2u_{I-3}^n - u_{I-4}^n \right) + \Delta t f_{I-2}^n \quad \{LW\}$$

4 Resultados e Discussão

Para ilustrar o caráter dispersivo da derivada u_{xxx} , começamos mostrando, na Figura 1, o comportamento da solução analítica ao longo do tempo. Vale destacar que, não só a solução analítica, como todas as simulações numéricas utilizaram uma base de partição do intervalo $[0, 1]$ com 100 pontos, isto é, $\Delta x = 0,01$. Foi escolhido, a princípio, uma relação de $\Delta t = \Delta x^3$, ou seja, $\Delta t = 0,000001$.

A solução numérica obtida a partir do método de Euler explícito – central apresentou boa concordância com a solução exata até 10 iterações, como mostra a Figura 2(a). A partir daí, uma instabilidade se inicia nas proximidades da fronteira $x = 1$ e com 15 iterações essa instabilidade já é bem pronunciada na região. Nesse instante, pode-se verificar também, a formação de instabilidade nas proximidades da fronteira $x = 0$. Essas instabilidades estão mostradas na Figura 2(c). Com 100 iterações, percebe-se que ambas instabilidades evoluíram e o erro global já se tornou extremamente grande. No entanto, há uma tendência da instabilidade

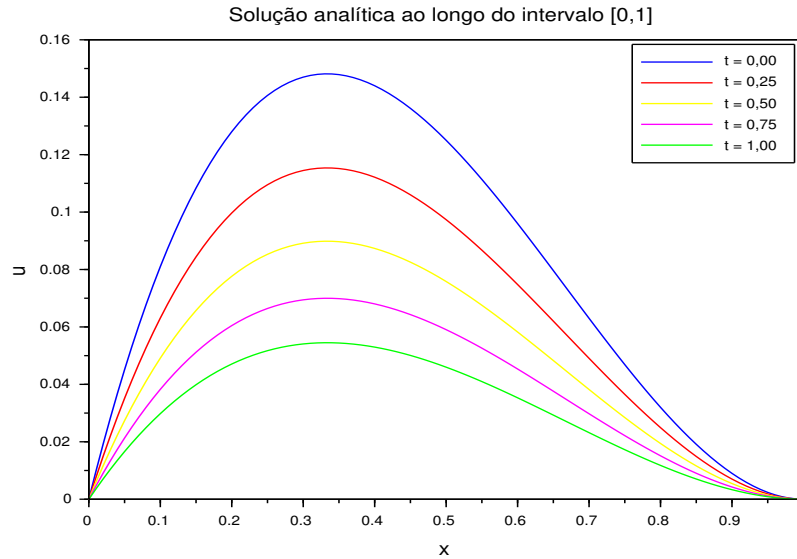


Figura 1: Solução analítica em instantes de tempo distintos.

oriunda de $x = 1$ prevalecer e se sobrepujar em relação a instabilidade de $x = 0$, assim como observado na Figura 2(e).

O método de Euler explícito – Holden concorda bem com a solução exata só até a quinta iteração, a partir daí, assim como ocorreu com EE-C, aparecem instabilidades em $x = 1$ e $x = 0$, respectivamente, como mostrado nas Figuras 2(b) e 2(d). O método EE-H apresenta um crescimento do erro global maior ainda do que EE-C, pois para um mesmo número de iterações ($n = 100$) seu erro é de uma ordem de grandeza mais alta que EE-C, como pode ser visto na Figura 2(f). Além disso, nesse caso, pode-se observar a hegemonia total da instabilidade oriunda de $x = 1$ frente a da outra extremidade. De certa forma, verifica-se o transporte da instabilidade de $x = 1$ no sentido de decrescimento do eixo x .

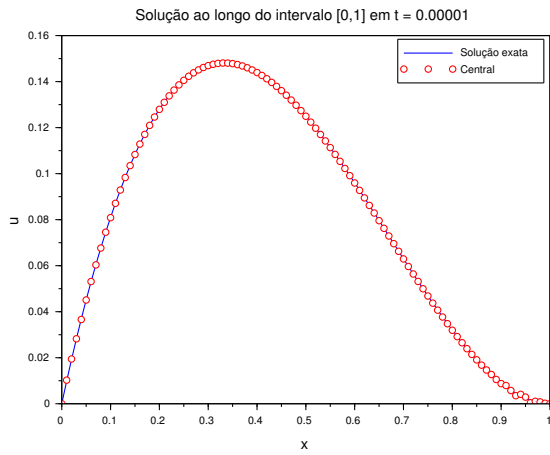
Já o método de Lax-Friedrichs, apresenta um comportamento inicial similar ao EE-H, porém, para 100 iterações, possui um erro na mesma ordem de grandeza do erro de EE-C (Figuras 3 e 4(a)). Todavia, diferente de EE-C, com 100 iterações, a instabilidade vinda do bordo $x = 0$ é bem mais contida, como mostra a Figura 4(b).

Finalmente, o método de Lax-Wendroff dá sinais de instabilidades com pouquíssimas iterações e seu erro evolui abruptamente, atingindo valores absolutos altíssimos, como pode ser observado nas Figuras 4(c) e 4(d).

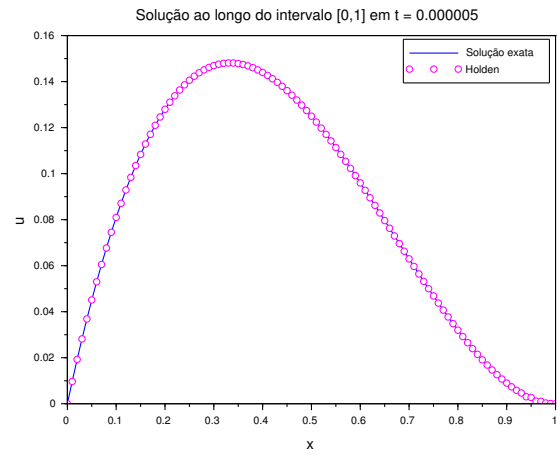
5 Conclusão

A principal limitação pertinente aos quatro métodos explícitos é a dependência de uma malha muito refinada para obtenção de bons resultados de integração. Todos os métodos oferecem uma margem muito estreita entre o instante inicial e o instante em que instabilidades começam a surgir. Ou seja, por mais refinada que seja a malha, é bem provável que o erro de todos os métodos comece a se descontrolar bem antes de $t = 1$. Uma possível solução a ser experimentada é a utilização de métodos implícitos em Diferenças Finitas.

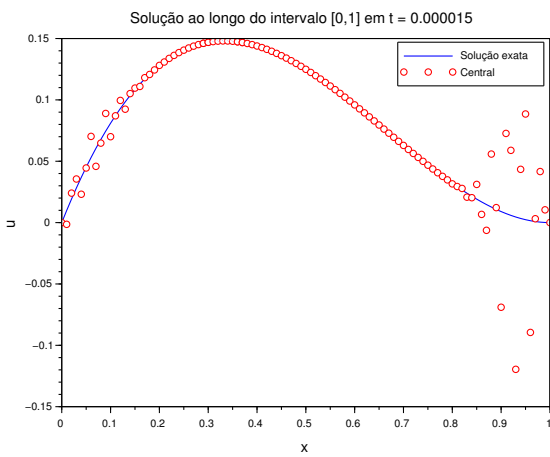
Outra fator, que agrava mais ainda o primeiro, é a grande influência que as condições de contorno – sobretudo as condições de contorno impostas em $x = 1$ – exercem sobre os demais pontos da malha. As instabilidades surgem nas fronteiras e se propagam “para dentro” do domínio da solução. Nesse sentido, deve-se procurar maneiras melhores de realizar o tratamento das condições de contorno de forma a reduzir essa interferência.



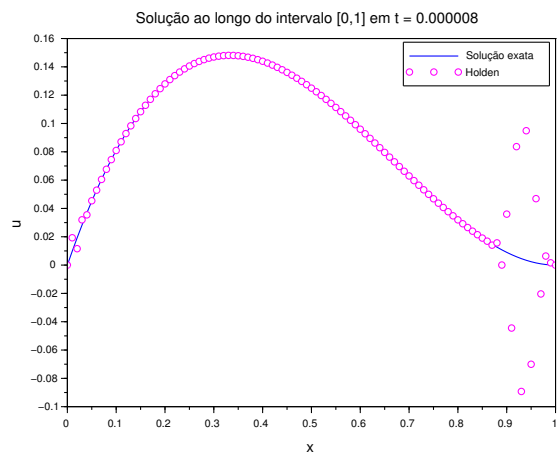
(a) EE-C



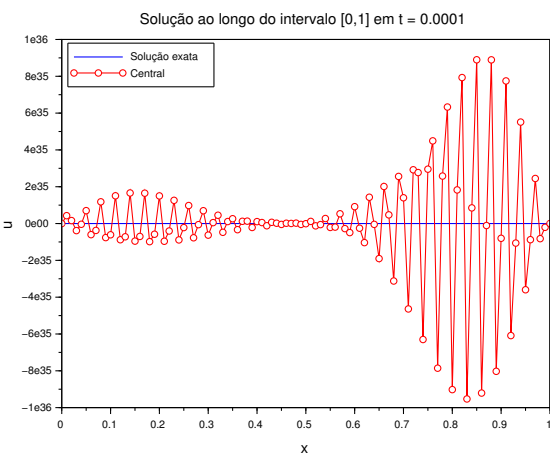
(b) EE-H



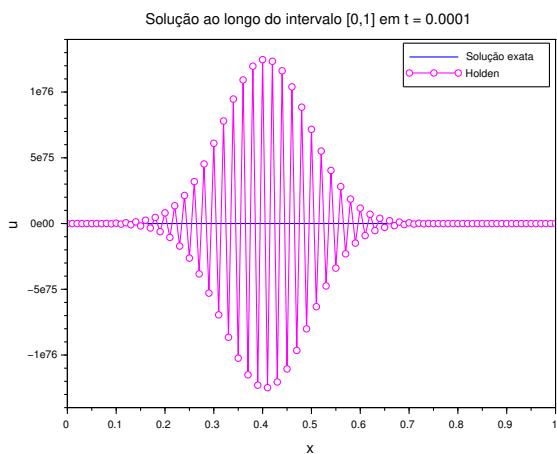
(c) EE-C



(d) EE-H



(e) EE-C



(f) EE-H

Figura 2: (a) e (b) Boa concordância exata dos métodos de EE-C e EE-H com a solução exata. (b) e (c) Formação de instabilidades nas extremidades do intervalo. (e) e (f) Predominância da instabilidade oriunda de $x = 1$.

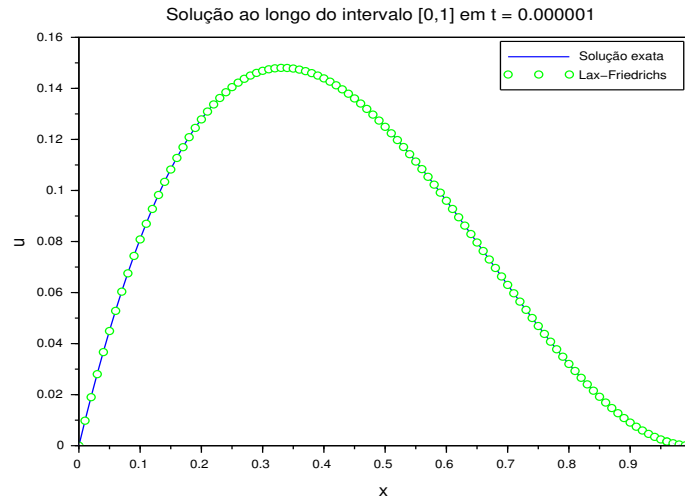
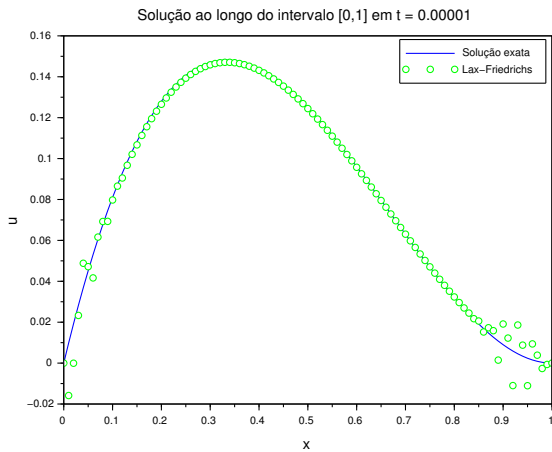
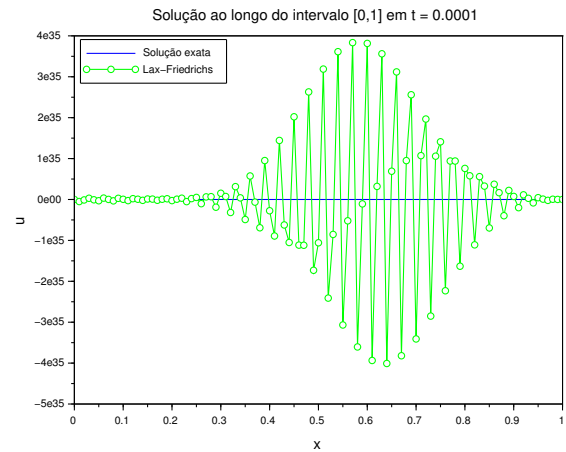


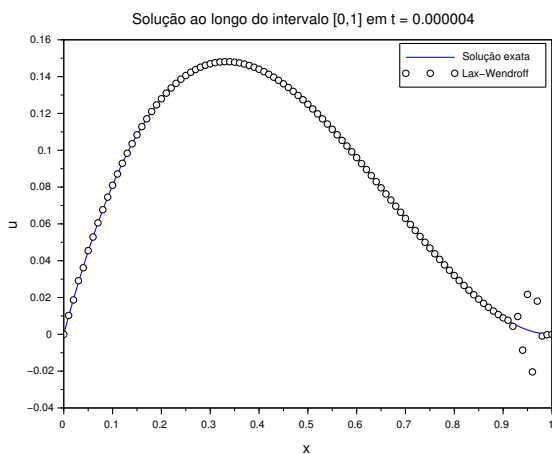
Figura 3: Boa concordância do método de LF com a solução exata.



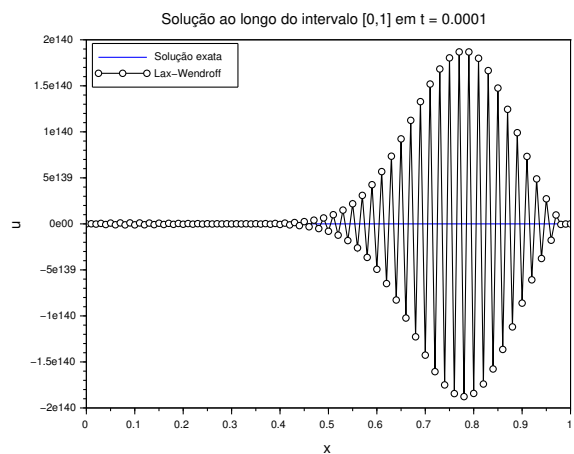
(a) LF



(b) LF



(c) LW



(d) LW

Figura 4: (a) e (c) Formação de instabilidades proveniente das fronteiras. (b) e (d) Evolução das instabilidades.

Referências

- [1] M. J. Ablowitz. *Nonlinear Dispersive Waves: Asymptotic Analysis and Solitons*. Cambridge University Press, Cambridge, 2011.
- [2] F. A. C. C. Chalub, J. P. Zubelli. Sólitos: Na Crista da Onda por mais de 100 anos., *Matemática Universitária*, 30:41-65, 2001.
- [3] C. Constantino, Ondas não-lineares em meio dissipativos descritas pela equação de Korteweg–de Vries–Burgers, Dissertação de Mestrado em Física, Unicamp, (1979).
- [4] J. A. Cuminato, M. Meneguetti Junior. *Discretização de Equações Diferenciais Parciais: Técnicas de Diferenças Finitas*. SBM, 2013.
- [5] H. Holden, U. Koley, N. H. Risebro. Convergence of a fully discrete finite difference scheme for the Korteweg–de Vries equation, *IMA Journal of Numerical Analysis*, 35:1047–1077, 2015. DOI: 10.1093/imanum/dru040.
- [6] D. J. Korteweg, G. de Vries. On the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary, *Philosophical Magazine Series 5*, 39:422-443, 1895. DOI: 10.1080/14786449508620739.
- [7] R. J. LeVeque. *Finite Difference Methods of Ordinary and Partial Differential Equations – Steady-State and Time-Dependent Problems*. SIAM, Philadelphia, 2007.
- [8] J. C. Strikwerda. *Finite Difference Schemes and Partial Differential Equations*. SIAM, Philadelphia, 2004.
- [9] F. S. Teixeira, Análise e Simulação Numérica da Equação de Korteweg-de Vries, Dissertação de Mestrado em Informática, UFRJ, (2011).
- [10] C. F. Vasconcellos, P. N. da Silva. Stabilization of the Kawahara equation with localized damping, *ESAIM: Control, Optimisation and Calculus of Variations*, 17:102-116, 2011. DOI: 10.1051/cocv/2009041.

O uso do Microsoft Kinect para manipular objetos

3D no CAVE

Iniciação Científica

Bolsista

Jean Rosa Manhães

Orientador

Jauvane C. de Oliveira

Petrópolis

Julho de 2018

1. Introdução

A proposta desse projeto foi desenvolver um controlador que pudesse substituir o uso do mouse 3D utilizado para interagir com o CAVE ou caverna digital (Cave Automatic Virtual Environment). A caverna digital é importante para diversos trabalhos como por exemplo trabalhos exploratórios, a universidade politécnica de São Paulo utiliza para tomadas de decisões críticas na área da engenharia [7].

Com a demanda das apresentações dos modelos 3D no CAVE foi possível notar a importância de um controlador que fosse prático e eficiente. O controlador por voz e gesto foi desenvolvido com o objetivo de otimizar a manipulação dos modelos 3D, fornecer autonomia e independência ao usuário uma vez que basta utilizar comandos de gesto ou voz para controlar o modelo.

2. Objetivo

2.1 Objetivo Geral

- 1) Desenvolver um controlador que fosse capaz de manipular por gesto e voz os modelos que são executados no instant reality e exibidos no cave.

2.2 Objetivos Específicos

- 1) Desenvolver uma conexão udp para transportar os dados do controlador para o instant reality;
- 2) Desenvolver bibliotecas para facilitar e otimizar o uso do servidor udp em projetos futuros;
- 3) Desenvolver um controlador que possa ser implementado e utilizado em diversos dispositivos;

3. Metodologia

Para o desenvolvimento do controlador foi necessário antes estudar as funcionalidades de todos os elementos que eram necessários para o projeto, analisar conceitos detalhados do kinect, quais linguagens de programação atenderiam o propósito, como seria feita a conexão e comunicação do controlador com o CAVE, como é gerado o código x3d do modelo 3D e por fim qual ferramenta é utilizada para exibir os modelos 3D no CAVE.

3.1 Kinect

Em 2017 foi realizado um estudo sobre o kinect pela plataforma da microsoft para entender como utilizá-lo no computador, quais ferramentas e bibliotecas precisariam baixar e como funciona cada sensor do kinect. Segundo a Microsoft [1], para utilizar o kinect v1 é necessário ter o sistema Windows 7 ou Windows 8, processador 32-bit (x86) ou 64-bit (x64), processador Dual-core, 1.66-GHz ou superior, USB 2.0 dedicado para o sensor do kinect v1, 2 GB de memória RAM, placa gráfica com suporte ao DirectX 9.0c.

“Em fevereiro de 2011, microsoft anunciou o lançamento do SDK (Software Development Kit), que pode ser obtido de forma gratuita e que permitiria aos desenvolvedores criar aplicativos utilizando as linguagens C++, C# e Visual Basic” [2].

3.2 Kinect na Unity

Foi feito um estudo de programação para kinect pelo visual studio [2], onde foi possível compreender parte do comportamento do kinect e como aplicá-lo no projeto. Em seguida seria necessário adaptar todo o conhecimento aprendido e implementá-lo utilizando a engine unity onde foi desenvolvido o controlador com kinect utilizando gesto e voz. Foi necessário consultar a documentação do kinect para unity [3], para entender as chamadas dos sensores de voz e gesto dentro da unity, também foi utilizado um pacote K2 (Kinect v2) disponibilizado por Filkov (2018) de forma gratuita em uso acadêmico e pesquisas.

3.3 Servidor UDP

Para a elaboração do servidor udp foi estudado a estrutura de comunicação e transporte de dados pela documentação da unity [4]. Após entender sua estrutura utilizou-se a linguagem *c#* e a ferramenta visual studio onde foi possível desenvolver tanto o cliente quanto o servidor udp, porém com a finalidade de otimizar a comunicação de futuros projetos também foram desenvolvidas duas dlls para o servidor e o cliente onde facilita e otimiza a implementação e o uso da comunicação udp em futuros trabalhos e projetos.

3.4 Instant Reality

Fez-se necessário o estudo da documentação do Instant Reality para que fosse possível entender seu funcionamento e linguagens suportadas [5]. Para a comunicação do servidor com o modelo foi identificado duas alternativas, comunicar diretamente com o instant reality através de plugins feitos em *c++* ou comunicar diretamente com a codificação *x3d* do modelo 3D. Para a comunicação com o *x3d* é possível utilizar a linguagem C, C++, C#, Java, JavaScript e Python.

“O X3D é um formato de arquivo de padrões abertos isento de royalties e ratificado pela ISO, além de uma arquitetura de tempo de execução para representar e comunicar cenas e objetos em 3D” [6].

X3D Graphics Standards: Specification Relationships

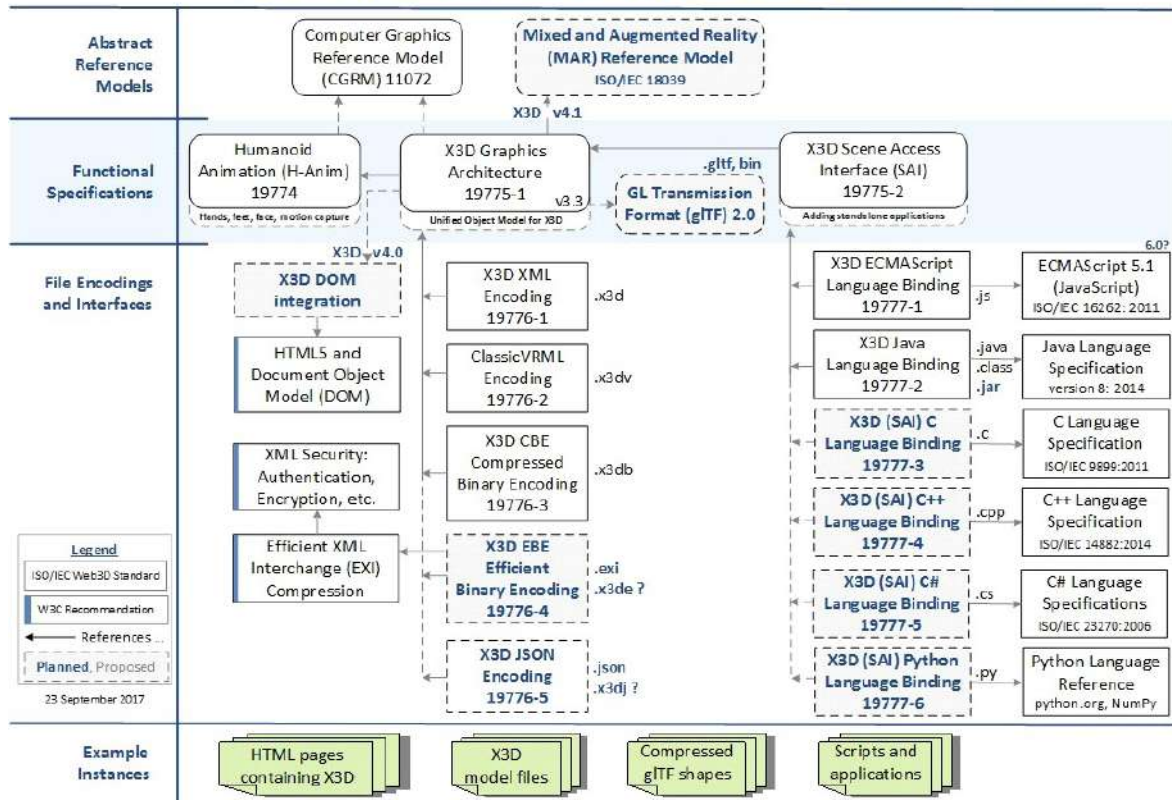


Figura 1- Padrões gráficos x3d: Especificação de relacionamentos [6].

4. Resultado

Identifica-se pelo menos 2 formas eficientes de se comunicar com os modelos 3D exibidos pelo CAVE, através de plugin feito em c++ ou comunicando o servidor UDP diretamente na codificação X3D do modelo. Escolhendo a segunda opção basta fazer a chamada do evento browser dentro da codificação c# ou a linguagem que optar de acordo com seu suporte [6] para realizar comunicação com o Instant Player e em seguida chamar os eventos do modelo X3D que deseja manipular conforme mostrado na documentação [5].

Há uma terceira forma de realizar a comunicação com o modelo fazendo com que o servidor UDP gere arquivos xml ou txt dos dados recebidos do controlador e em seguida utilizar alguma linguagem suportada [6] para ler esse arquivo e manipular o

código X3D. Porém não é aconselhável nem eficiente uma vez que é necessário fazer a conexão com o browser do Instant Player para atualizar os eventos do modelo. Ao manipular diretamente com o arquivo XML o modelo apenas sofrerá modificação após ser recarregado, pois não há um evento de update disponível.

Percebeu-se que há limitações de dados a serem transferidos para o modelo, a codificação suporta uma determinada quantidade de chamadas de eventos, para realizar grandes quantidades de chamadas faz-se necessário utilizar o evento de update do browser [5].

5. Conclusão

Conclui-se que o uso do kinect para manipular objetos do Instant Reality é de grande eficiência e praticidade, porém por utilizar o modelo v1 do kinect percebeu-se algumas falhas em sua execução por tempo prolongado.

Faz-se necessário a substituição para o modelo v2 do kinect para correções de falhas e melhorias no controlador.

5.1 Execução

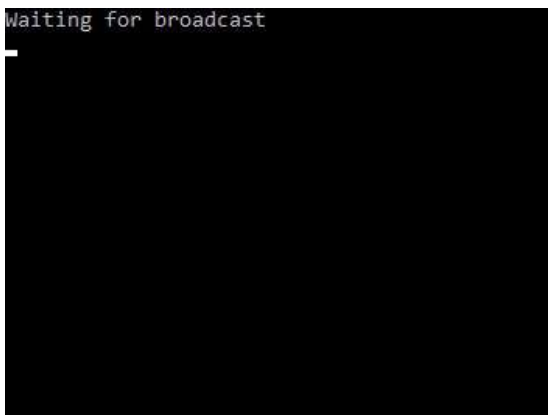


Figura 2 – Servidor Udp



Figura 3 – Iniciando o Controlador Kinect e passando o endereço de IP do servidor UDP



Figura 4 – Controlador Kinect

Referência Bibliográfica

- [1] Microsoft Virtual Academy. Crie seu jogo para kinect. Disponível em: <https://mva.microsoft.com/pt-br/training-courses/crie-seu-jogo-para-kinect-14184?l=Us51eY4eB_2305513172> . Acessado em: 10 de fev. 2018.
- [2] SCHADE, Gabriel. Microsoft Kinect, crie aplicações interativas. São Paulo, p.3-4, 2017.
- [3] K2. Documentação Kinect-Unity. Disponível em: <<https://ratemt.com/k2docs/ModelFaceController.html>>. Acessado em: 22 de fev. 2018.
- [4] Unity. Documentação. Disponível em: <<https://docs.unity3d.com/Manual/UNetUsingTransport.html>>. Acessado em: 10 de fev. 2018.
- [5] Instant Reality . Documentação. Disponível em: <<http://doc.instantreality.org>>. Acessado em : 12 de jul. 2018.
- [6] Web3D. X3D Graphics standards: Specification Relationships. Disponível em: <www.web3d.org>. Acessado em: 20 de jun. 2018.
- [7] Escola Politécnica, Universidade de São Paulo. Caverna Digital, quebrando paradigmas. Disponível em: <<http://www.poli.usp.br/fr/comunicacao/noticias/destaques/arquivo-em-foco/368-caverna-digital-quebrando-paradigmas.html>>. Acessado em: 10 de jul. 2018.

RELATÓRIO DE ATIVIDADES

DADOS GERAIS

Título do Projeto: *Gerência e Análise de Dados Científicos de Workflows de Filogenia Aplicada a Genomas de Arbovírus.*

Nome do Bolsista: José Maria de Souza Junior

Nome do Orientador(a): Kary Ann del Carmen Ocaña Gautherot

Nome do Coorientador: André Elias Rodrigues Soares e Luiz Manoel Rocha Gadelha

Tipo de Bolsa: PIBIC

Período do Relatório: 01/03/2018 – 10/07/2018

OBJETIVOS

Esse projeto de iniciação científica tem como objetivo propor uma série de soluções para a gerência e armazenamento de simulações computacionais e dados científicos de bioinformática, por meio de tecnologias especializadas de sistemas de gerência de *workflows* científicos e banco de dados. Será implementado um esquema de banco de dados para dados de proveniência gerados de um *workflow* científico filogenético e acoplado a infraestrutura de processamento de alto desempenho no LNCC. Finalmente, o arcabouço desenvolvido será validado com dados reais de estudos filogenéticos de múltiplos genomas de vírus.

INTRODUÇÃO

Explorar e analisar a grande quantidade de informação biológica é atualmente um desafio envolvendo áreas como Ciência da Computação, Biologia Computacional [Marx 2013] e Ciências de Dados. Grandes volumes de dados normalmente ocultam padrões interessantes e úteis que não são identificados trivialmente. A análise manual desses dados pode ser inviável, e mesmo que seja realizada, é passível de erros e pode não extrair todo o conhecimento existente. Dessa forma, tecnologias de banco de dados, Sistemas de Gerência de *Workflows* Científicos (SGWfC) e aprendizado de máquina no apoio a análise de dados de bioinformática tornam-se necessários, podendo apoiar a descoberta de padrões de sequências, funções de genes, interações proteína-proteína ou redes de filogenia que alavanquem investigações na terapia de doenças. Desta maneira, explorar a informação dos dados de proveniência de forma estruturada e armazenar essa informação em um banco de dados persistente é necessário a fim de poder acessá-la para inferir conhecimento [Kashyap H et al. 2015; Ocaña et al. 2015].

Workflows científicos se tornam mais complexos ao manipularem dados científicos volumosos, heterogêneos ou distribuídos [Mattoso et al. 2010], pelo que precisam de ambientes e tecnologias de processamento de alto desempenho (PAD), SGWfC e de banco de dados. Esses dados de proveniência [Freire et al. 2008] envolvem metadados de execução, gerência, acesso e armazenamento que contém informações específicas de um experimento científico, como o seu ciclo de vida, atividades e programas, e desempenho do *workflow*. Além disso, tecnologias como algoritmos de mineração de dados, aprendizado de máquinas são naturalmente acopladas em pesquisas de banco de dados e bioinformática, o que não é uma tarefa simples.

O projeto atual se enquadra nas pesquisas do projeto RNBio¹ e nos três projetos em andamento sob coordenação da orientadora: JCNE² (FAPERJ), Grupos Emergentes³ (FAPERJ) e Universal⁴ (CNPq). A motivação desse projeto reside em explorar os benefícios de tecnologias de banco de dados e *workflows* aplicados nas áreas de pesquisa de bioinformática e levá-los aos ambientes de PAD, aliados ao uso de ferramentas complementares que estão sendo desenvolvidas no LABINFO/LNCC e colaboradores. Cabe mencionar que esse projeto de IC dará o suporte computacional necessário às pesquisas de filogenia em larga escala e aos dois projetos de IC já em andamento também orientados pela Kary e André: (i) *workflow* filogenético desenvolvido pelo Gustavo Saboia e (ii) testes de escalabilidade de programas de filogenia no Santos Dumont pela Joice Pedro.

Na primeira fase do presente projeto foram exploradas duas ferramentas de análise de dados de proveniência: (i) noWorkflow, desenvolvido pela Universidade Federal Fluminense (UFF) e Universidade de Nova York (NYU), realiza a gerência de forma não intrusiva *i.e.*, desde o código original dos *scripts* (*e.g.*, Python) e a análise de dados de proveniência independente do uso (ou não) de SGWfC e (ii) SAMbA, desenvolvido pela UFF, é uma extensão do Apache Spark que gerencia a execução de programas na forma de caixa preta e coleta, armazena e consulta dados de proveniência com base nos rastreamentos de proveniência usando PAD.

METODOLOGIA

A primeira etapa do projeto consistiu no estudo de conceitos de gerência e análise de dados, *workflows* científicos, bioinformática, filogenia e tecnologias de PAD e da estrutura de um *workflow* científico de filogenia, suas dependências e funções.

Na segunda etapa foi explorado o noWorkflow [Murta et al. 2014] e foram realizadas execuções usando exemplos propostos pelos desenvolvedores e usando o estudo real do *workflow* de filogenia.

A terceira etapa (em andamento) envolve o acoplamento e uso eficiente do SAMbA. Foram realizados testes com os exemplos dos desenvolvedores e com

¹ Estruturação e implantação da Rede Nacional de Bioinformática, MCTI, CNPq. CNPq 456644/2013-0

² Apoio Computacional à Análise Genômica de DTN por meio de Mineração de Dados e Aprendizado de Máquina, JCNE_2016

³ Gerenciamento e Análise de Dados Biológicos em Plataformas de HPC e de Processamento de Big Data, PRONEM Nº 04/2016

⁴ Mineração de Dados de Experimentos de Bioinformática Executados em Larga Escala no Apoio ao Estudo de Doenças Negligenciadas e Doenças Genômicas em Humanos, Edital CNPq Universal 2016 - Processo No. 429328/2016-8

workflow de filogenia de maneira satisfatória, restando a implementação do SAMbA nos *clusters* do LNCC e testes de desempenho.

RESULTADOS E DISCUSSÃO

A primeira etapa do projeto abordou o levantamento bibliográfico e estudo de conceitos de bioinformática, filogenia e dados e *workflows* científicos e foi realizada com sucesso. Foi realizado um estudo da estrutura, classes e funções dos *scripts* que compõem o *workflow* científico desenvolvido pelo Gustavo, dado que ele será usado como caso de estudo desse projeto.

Como resultado desse estudo, podemos afirmar que esse *workflow* de filogenia é um bom candidato para ser usado no presente projeto proposto, o que permitirá reforçar soluções de gerência de execução de *workflows* em larga escala e de análise de dados de proveniência. Dessa maneira, dois SGWfC foram eleitos para esse fim, o noWorkflow e o SAMbA.

Na segunda etapa exploramos a bibliografia e estrutura do noWorkflow. Testes com exemplos dos desenvolvedores foram bem sucedidos, mas quando a integração foi feita ao *workflow* de filogenia surgiram vários problemas e erros. Provavelmente relacionados à estrutura conceitual do noWorkflow que não suporta o a implementação com *workflows* muito mais complexos, característicos das áreas como a bioinformática. Por tanto se tornou necessário uma ferramenta mais robusta e que permita o paralelismo e distribuição de tarefas.

A Figura 1 apresenta o resultado da execução bem-sucedida de um exemplos do noWorkflow utilizando a classe Simulator como módulo de execução. O gráfico é a representação visual da execução das etapas do projeto (representado por círculos) e as interconexões entre as dependências (linhas).

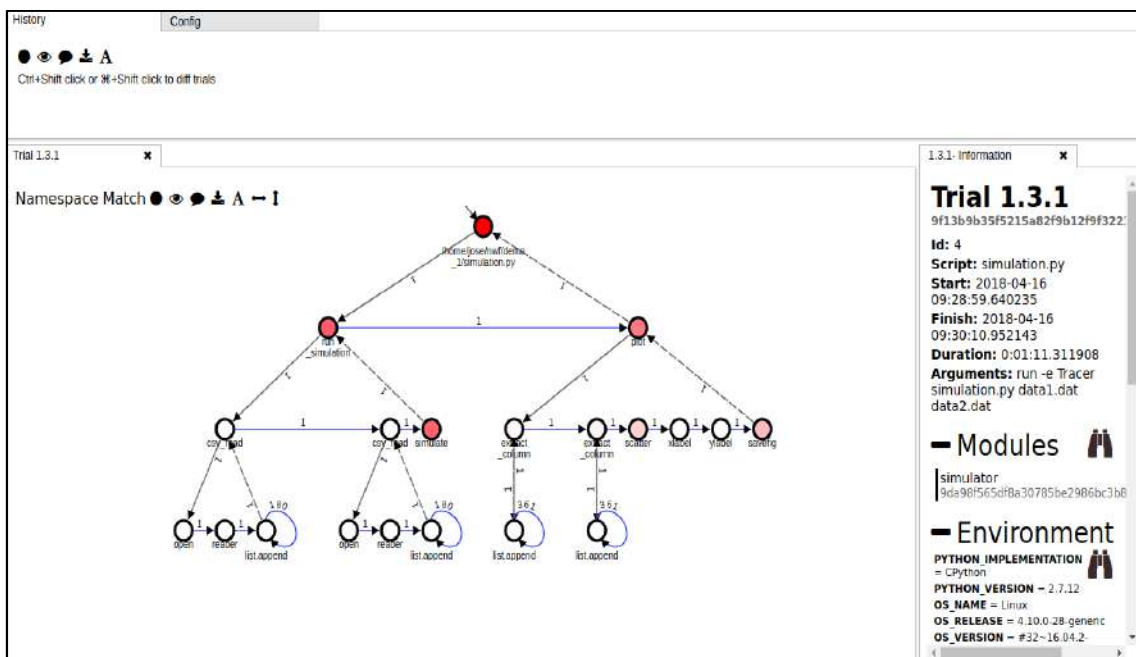


Figura 1. Resultado de uma execução do noWorkflow com o exemplo proposto

A Figura 2 apresenta o resultado gerado pelo noWorkflow usando o *workflow* de filogenia. Em comparação com a Figura 1, não foram geradas a maioria das saídas da execução ou as suas interligações de dependências. A informação que foi extraída e armazenada pelo noWorkflow foi relacionada a arquivo inicial e dos *scripts* executados, mas nenhuma outra informação extraída das funções (do *script*) relacionada a metadados foi gerada. Foi demonstrado que o noWorkflow não suporta as requisições necessárias de gerenciamento de metadados do nosso *workflow*.

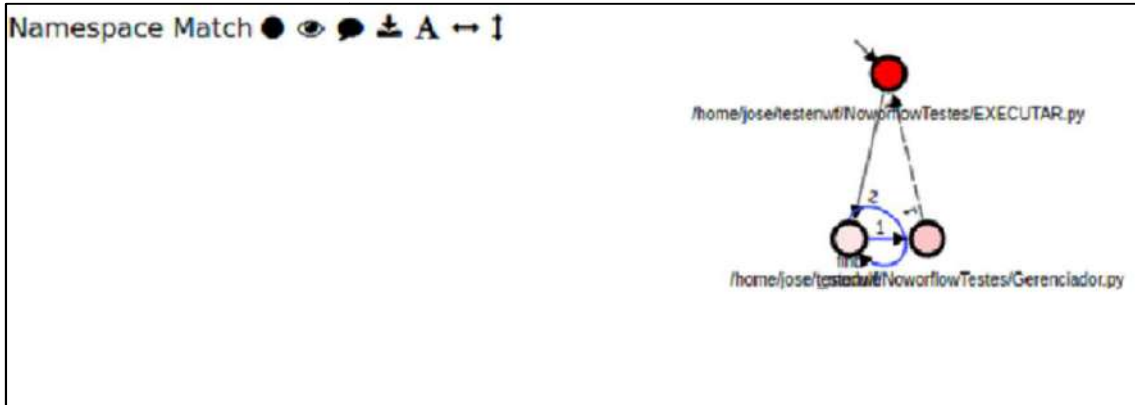


Figura 2. Resultado de uma execução do noWorkflow com o *workflow* de filogenia

Na terceira etapa exploramos a bibliografia e estrutura do SAMbA, incluindo o estudo da linguagem de programação Scala e o banco de dados Cassandra usados pelo SAMbA. Testes com exemplos e com o *workflow* de filogenia SciPhy [Ocaña et al. 2011] foram bem sucedidos. Os resultados dos elementos usados para armazenar dados de proveniência do *workflow* foram gerados via interface *Web*. A Figura 3 mostra a execução do SciPhy usando o arquivo de entrada ORTHOMCL888, as atividades do *workflow* (*readseq*, *modelgenerator*, *filegroup*, *mafft* e *raxml*) e os dados de proveniência que foram extraídos e armazenados no banco de proveniência usando o Cassandra. Esses resultados permitiram demonstrar que o SAMbA se adéqua as características de um *workflow* de filogenia pelo qual é o SGWfC eleito nesse projeto.

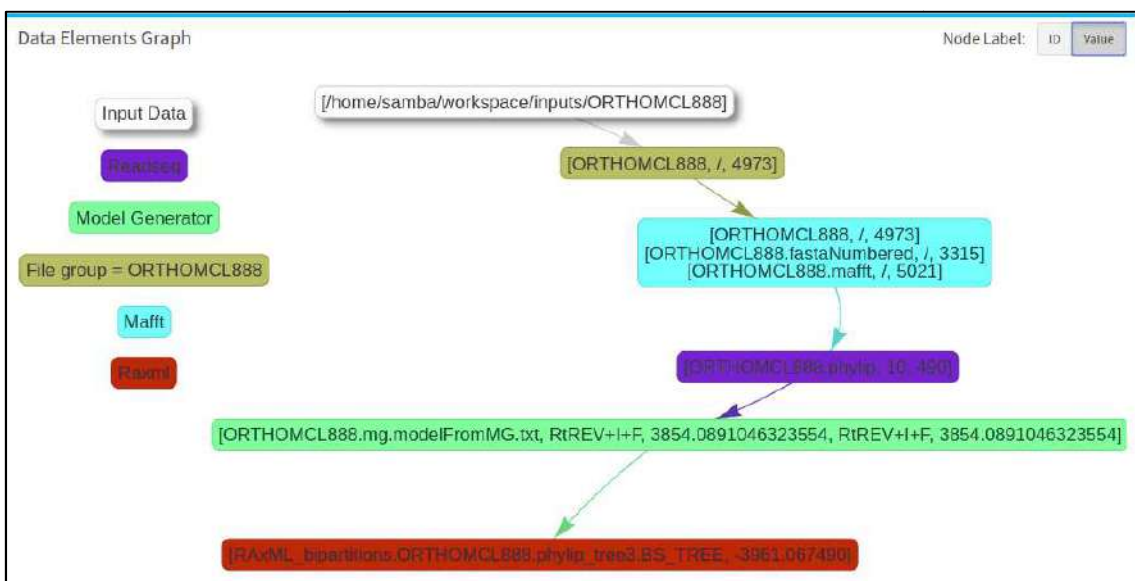


Figura 3. Resultado de uma execução do SAMbA com o *workflow* de filogenia

CONCLUSÃO

A análise de experimentos científicos envolvendo dados biológicos é complexa e requer estudos, ferramentas e infraestrutura robusta. A utilização do SAMbA mostrou ser eficiente. Os próximos passos nesse projeto são a integração do SAMbA com o *workflow* filogenético e a execução no ambiente de PAD do LNCC. Com isso teremos o resultado funcional do *workflow* desenvolvido em conjunto com o SGWfC proposto rodando de forma distribuída.

REFERÊNCIAS BIBLIOGRÁFICAS

Freire, J., Koop, D. and Moreau, L. [Eds.] (2008). *Provenance and Annotation of Data and Processes*. Berlin, Heidelberg: Springer Berlin Heidelberg. v. 5272

Kashyap H, Ahmed HA, Hoque N, Roy S and Bhattacharyya DK. (2015). Big Data Analytics in Bioinformatics: A Machine Learning Perspective. *Computing Research Repository*,

Marx, V. (13 jun 2013). Biology: The big challenges of big data. *Nature*, v. 498, n. 7453, p. 255–260.

Mattoso, M., Werner, C., Travassos, G. H., et al. (2010). Towards supporting the life cycle of large scale scientific experiments. *International Journal of Business Process Integration and Management*, v. 5, n. 1, p. 79–92.

Murta, L., Braganholo, V., Chirigati, F., Koop, D. and Freire, J. (2014). noWorkflow: Capturing and Analyzing Provenance of Scripts. In *International Workshop on Provenance Annotation (IPAW)*.

Ocaña, K., De Oliveira, D., Silva, V. and Mattoso, M. (2015). Data Analytics in Bioinformatics: Data Science in Practice for Genomics Analysis Workflows. In *Proceeding of the 11th IEEE International Conference on eScience*.

Ocaña, K., Oliveira, D. De, Ogasawara, E., et al. (2011). SciPhy: A Cloud-Based Workflow for Phylogenetic Analysis of Drug Targets in Protozoan Genomes. In *Advances in Bioinformatics and Computational Biology*. , Lecture Notes in Computer Science. Springer. http://link.springer.com/chapter/10.1007/978-3-642-22825-4_9.

Relatório Iniciação Científica

18 de Julho de 2018

1 Dados Gerais

Título: Modelagem e Controle de um Manipulador Robótico Sujeito a Falhas

Bolsista: Junior do Nascimento Xavier

Orientador: Marcos Garcia Todorov

Bolsa: PIBIC - 01/03/2018 até 31/07/2018

2 Objetivos

- 2.1 Integração numérica (Método de Euler) e sua aplicação no software Matlab.
- 2.2 Estudo dirigido em Álgebra Linear.
- 2.3 Estudo do modelo que descreve o manipulador robótico.
- 2.4 Estudo dirigido em Probabilidade e Processos Estocásticos.
- 2.5 Pôster para a Jornada de Iniciação Científica.

3 Introdução

Atualmente há diversos sistemas de manipuladores robóticos aplicados nas mais variadas áreas, tais como: aplicações nos processos industriais, aplicações em medicina, entre outras. Como quaisquer outros mecanismos, estes não estão livres de falhas de atuação, sejam tais falhas oriundas dos seus operadores, de fatores ambientais ou outros fatores possíveis de causá-las. O objetivo desta pesquisa foi compreender inicialmente o complexo modelo matemático que rege o manipulador robótico, e analisar as falhas de atuação que podem ocorrer devido à sua natureza. O mecanismo subatuado (quando o número de atuadores é menor que o número de graus de liberdade), permite algumas simulações de falhas.

4 Materiais e Métodos

Inicialmente o plano de estudo era compreender como funciona a integração numérica via método de Euler, a compreensão do mesmo era importante, pois o ambiente utilizado para simular as falhas de atuação, utiliza integração numérica via método de Euler para realizá-las. Para execução da pesquisa principal, foi necessário estudar sobre o modelo que governa o manipulador robótico. Para isso a pesquisa baseou-se no livro *Robust Control of Robots* [1] e na tese de doutorado *Controle H_∞ não linear de robôs manipuladores subatuados* [2]. As duas fontes de pesquisa foram de extrema importância, outros artigos sobre manipuladores robóticos também foram utilizados.

O livro possui um ambiente baseado em Matlab, que permite a simulação de tais falhas, onde é possível colher dados graficamente das simulações. Tal ambiente permite simular diversas falhas de atuação, além de permitir ajustar configurações desejadas e especificidades do manipulador, tais como: inércia, gravidade, fricção, entre outras. O livro e a tese, abordam a modelagem do manipulador robótico, além de explicarem teoricamente como ocorrem as falhas de atuação.

5 Resultados e Discussões

As tarefas que foram designadas, foram concluídas de acordo com o previsto, entretanto, houveram alguns atrasos da parte do orientado. Os resultados obtidos foram os esperados, por exemplo, o aprendizado da modelagem do manipulador e suas falhas possíveis de ocorrer. Um estudo dirigido foi concluído e outro segue em andamento. Por fim, está sendo confeccionado um pôster para a Jornada de Iniciação Científica, dentro da ideia da modelagem e das falhas do manipulador robótico.

6 Conclusões

Ao longo desses quatro meses de pesquisa, a ideia principal foi compreender a modelagem do manipulador e as falhas que podem ocorrer, o que foi de extrema importância, pois possibilitará não só a apresentação do pôster na Jornada de Iniciação Científica, como um aprofundamento do assunto no futuro. Para o futuro, o que se planeja é ir mais a fundo nas questões das falhas, cujo o objetivo será o possível tratamento das falhas que ocorrerem, visando o melhor desempenho para o sistema. .

7 Referências Bibliográficas

Referências

- [1] Adriano A. G. Siqueira, Marco H. Terra, Marcel Bergeman. *Robust Control of Robots: Fault Tolerant Approaches*. Springer-Verlag, London, 2011.
- [2] Adriano A. G. Siqueira. Controle H_∞ não linear de robôs manipuladores subatuados. Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2004.

Treinamento para comissários de bordo

Iniciação Científica

Bolsista

Lenon Baldez Stelman

Orientador

Jauvane C. de Oliveira

tipo de bolsa – Iniciação Científica
período do relatório – Outubro de 2017 / Julho 2018

Petrópolis
Julho de 2018

1 - Introdução

A proposta desse projeto foi desenvolver um simulador para treinamento de comissários de bordo.

O simulador será multiplayer, podendo jogar mais de uma pessoa ao mesmo tempo, será dividido em dois ambientes, o ambiente para o treinador, que criará situações para o comissário aprendiz e o ambiente para o comissário aprendiz realizar as tarefas forem pedidas.

Sendo assim, o comissário de bordo terá que analisar as ações dos passageiros e identificar o que eles têm, os passageiros podem ter enjôos, podem estar querendo água ou até mesmo ficar em pânico.

2 - Objetivo

Objetivo Geral

- 1) Desenvolver um game simulador, que será capaz de treinar e capacitar o comissário de bordo [3], para reagir em diferentes situações, como por exemplo, o passageiro tem medo de viajar, o comissário terá que lidar com essa situação da melhor forma possível.

Objetivos Específicos

- 1) Desenvolver uma conexão multiplayer.
- 2) Modelar e criar o protótipo 3D do avião e passageiros;
- 3) Desenvolver um banco de perguntas e respostas para as reações dos passageiros;

3 - Metodologia

Para o desenvolvimento do simulador foi necessário antes estudar as normas da ANAC [2], Agência Nacional de Aviação Civil é uma agência reguladora federal cuja responsabilidade é supervisionar a atividade de aviação civil no Brasil [1], depois de conhecer as normas, trabalharemos na modelagem 3D e criação dos objetos, como a aeronave e sua tripulação de passageiros, com os assets prontos entraremos na codificação e logo depois o multiplayer, que será com o treinador e o comissário de bordo aprendiz.

4 – Resultado

Ainda não chegamos em um resultado, estamos trabalhando em um protótipo.

5 – Conclusão

Ainda não chegamos em uma conclusão, estamos trabalhando em um protótipo.

6 - Referência Bibliográfica

- [1] Profissional da aviação. Requisito para ser comissário de bordo. Disponível em: <
<http://www.anac.gov.br/assuntos/setor-regulado/profissionais-da-aviacao-civil/profissionais/comissario-de-voe> > . Acessado em: 10 de fev. 2018.
- [2] Regulamentação ANAC. Disponível em: <
<http://www.anac.gov.br/assuntos/setor-regulado/profissionais-da-aviacao-civil/arquivos/mca5811.pdf> > . Acessado em 13 de jul. 2018.
- [3] Treinamento para comissários. Disponível em: <
<http://vidadecomissario.a350.com.br/gol-inaugura-novo-centro-de-treinamento-os-atuais-e-futuros-comissarios-da-cia-poderao-treinar-nele/> > . Acessado em: 22 de fev. 2018.



Relatório de Projeto

Estudo e Aplicação de Técnicas de Visualização para o Problema da Progressão Tumoral

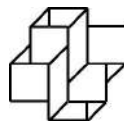
Bolsista:

Matheus Evangelista de Oliveira

Supervisoras:

Prof. Dr. Regina C. Almeida (LNCC)

Prof. Dr. Sicília Ferreira Ponce Pasini Judice (FAETERJ Petrópolis)



LNCC
PROGRAMA PIBIC/PIBITI
04/2018 - 07/2018

Julho de 2018

Resumo: Este relatório descreve as atividades desenvolvidas no contexto do Programa de Iniciação Científica do Laboratório Nacional de Computação Científica - LNCC/MCTI. O objetivo é capacitar e dar continuidade ao desenvolvimento de pesquisas relacionadas à modelagem computacional de crescimento tumoral. A carcinogênese é um processo complexo que normalmente

é afetado por muitos fatores em diferentes escalas tais como mudanças genéticas e fenotípicas, angiogênese e movimento celular. Modelos matemáticos e computacionais podem ajudar a entender, explicar e prever a carcinogênese e seu tratamento. Fundamental para este entendimento é a representação visual dos dados envolvidos.

Este projeto propõe o estudo e aplicações de técnicas de visualização científica para o problema da progressão tumoral. Especificamente o trabalho desenvolvido foi dar continuidade ao desenvolvimento de um visualizador que, a partir dos dados gerados pelo modelo híbrido para o crescimento avascular de tumores, gera uma visualização tridimensional. Os objetivos específicos foram:

- Otimização da interação do usuário com a ferramenta no que diz respeito ao controle e movimentação da câmera virtual.
- Exportação de imagem em alta qualidade.
- Armazenamento de sequência de imagens e gravação em formato de vídeo.

Este projeto foi desenvolvido no LNCC/MCTI, sob supervisão da Dr. Regina C. Almeida, pesquisadora titular e docente permanente do Programa de Pós-Graduação Multidisciplinar em Modelagem Computacional, com a colaboração da especialista em computação gráfica, Prof. Dr. Sicília Judice, da FAETERJ/Petrópolis.

1 Introdução

As estatísticas a respeito do câncer são alarmantes. Nos EUA, o risco de uma pessoa desenvolver câncer é de 16% (1 em 6), sendo 42% (1 em 2) em homens e 38% em mulheres (1 em 3) [1]. No Brasil, atualmente o câncer é a segunda maior causa de morte, com 190 mil óbitos por ano. Dados mais recentes apontam o câncer como a principal causa da mortalidade em 10% das cidades brasileiras [4]. Este cenário tem motivado o incremento das pesquisas em câncer. Modelos matemáticos e computacionais vêm sendo criados há bastante tempo para descrever o surgimento e crescimento de tumores. O Grupo de Modelagem do Crescimento Tumoral do LNCC tem também contribuído com o desenvolvimento de modelos contínuos, discretos e híbridos para representar uma série desses fenômenos multiescala característicos do câncer. A pesquisa desenvolvida resultou em uma tese de doutorado, defendida em 2014 [5], quatro dissertações de mestrado [2, 3, 7, 8], e dois trabalhos de tese de doutorado foram iniciados em 2016.

Uma das dissertações finalizadas recentemente apresenta um modelo híbrido para o crescimento avascular de tumores [7]. É um modelo multiescala híbrido bi-dimensional que integra fenômenos que ocorrem nas escalas celular e de tecido. A escala celular é descrita através de um modelo baseado em agentes, que possibilita tratar cada célula (normal e cancerosa) individualmente e descrever seu comportamento no microambiente. Cada célula é um agente que interage com os demais através de forças. O balanço de forças determina as novas posições das células através da segunda lei de Newton. Na escala do tecido representamos a dispersão de nutrientes (oxigênio) no meio através de uma equação diferencial parcial de reação-difusão. As células tumorais são diferenciadas por estados fenotípicos (quiescente, proliferativa, apoptótica, hipóxica e necrótica), que podem ser alterados em função das condições do meio. A dinâmica do crescimento tumoral é regida pelas transições entre os estados fenotípicos, algumas consideradas eventos determinísticos e outras como eventos estocásticos.

As simulações realizadas com o modelo híbrido desenvolvido, restritas à um número limitado (centenas) de células, demonstraram grande potencial para estudos mais realistas. Neste contexto, apesar do modelo desenvolvido ter sido desenvolvido de forma estruturada e flexível para a incorporação de novas dinâmicas e facilmente estendido para domínios tri-dimensionais, simulações mais realistas requerem uma estrutura computacional mais eficiente para representar e visualizar a dinâmica de milhares de células. Para facilitar esse processo é fundamental um ambiente visual que represente esses dados de forma clara e dinâmica para o usuário.

2 Metodologia

O desenvolvimento deste projeto foi feito com a linguagem de programação C++ em conjunto com a biblioteca de visualização OpenGL. Usou-se também a biblioteca OpenCV para tratar a exportação das imagens. O código fonte do projeto está disponível na plataforma GitHub através do endereço <https://github.com/tmglncc/Visualizador>.

O foco deste projeto é a visualização das células. Desta forma, todas as células têm uma morfologia circular, onde cada célula i tem os seguintes atributos: posição do centro (x_i), raio (R^i) e raio do núcleo (R_N^i). Uma célula pode interagir com outras células caso elas estejam dentro do raio de ação, denotado por R_A^i . A Figura 1(a) ilustra essa descrição geométrica. Células normais estão em homeostase e a heterogeneidade fenotípica nas células cancerosa é caracterizada pelos estados proliferativo (\mathcal{P}), quiescente (\mathcal{Q}), apoptótico (\mathcal{A}), hipóxico (\mathcal{H}), e necrótico (\mathcal{N}) ilustrados na Figura 1(b). As cores propostas nesse esquema de transição dos estados das células foram utilizadas no visualizador desenvolvido.

3 Resultados e Discussão

O visualizador desenvolvido recebe uma base de dados gerada pelo simulador e exibe um modelo bidimensional ou tridimensional, permitindo ao usuário manipular os dados de várias maneiras, como: alterando ângulos de visualização, definindo qual tipo de célula deve ser exibido, controlando os *frames* da simulação. As Figuras 2(a)-(b) ilustram a visualização de uma simulação

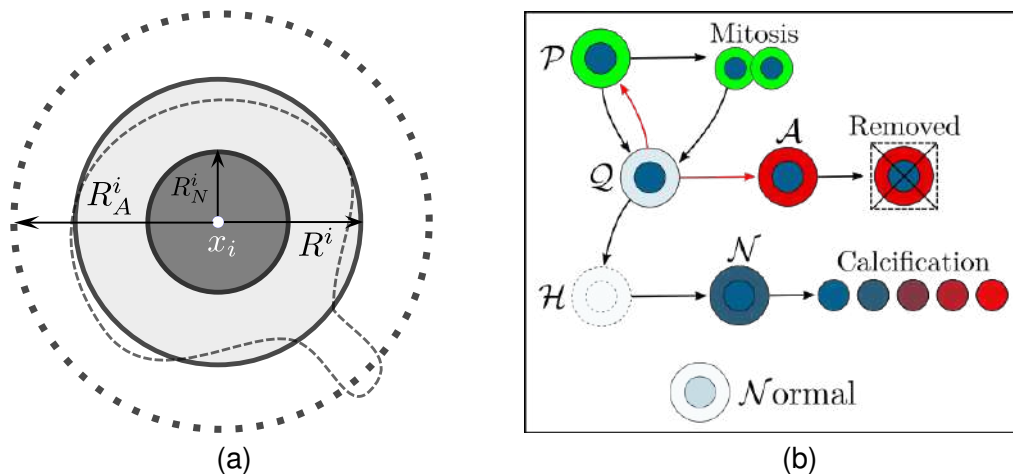


Figura 1: (a) Descrição geométrica da célula. (Fonte: [6].) (b) Esquema de transição das células.

bidimensional. Acionando teclas específicas no teclado, o usuário pode filtrar o tipo de célula que deve ser exibida, como mostra a Figura 2(b). As Figuras 2(c)-(d) ilustram a visualização de uma simulação tridimensional. Usando o *mouse* o usuário pode rotacionar o modelo e usando controle de teclas é possível realizar cortes no volume de visualização, como mostra a Figura 2(d).

Foi decidido que algumas alterações no visualizador precisavam ser feitas para que ele gerasse a imagem em alta resolução, e que cumprisse com o propósito de ser mais amigável ao usuário. Antes de executar essas alterações, a implementação do visualizador foi atualizada com o uso de um *software* de controle de versão Git, para garantir que todas as alterações pudessem ser revertidas e versões anteriores do visualizador pudessem ser recuperadas.

- Primeiramente a interação com o usuário deveria ser mais precisa e a visualização mais dinâmica em relação aos dados lidos pelo visualizador. Portanto foram realizados ajustes

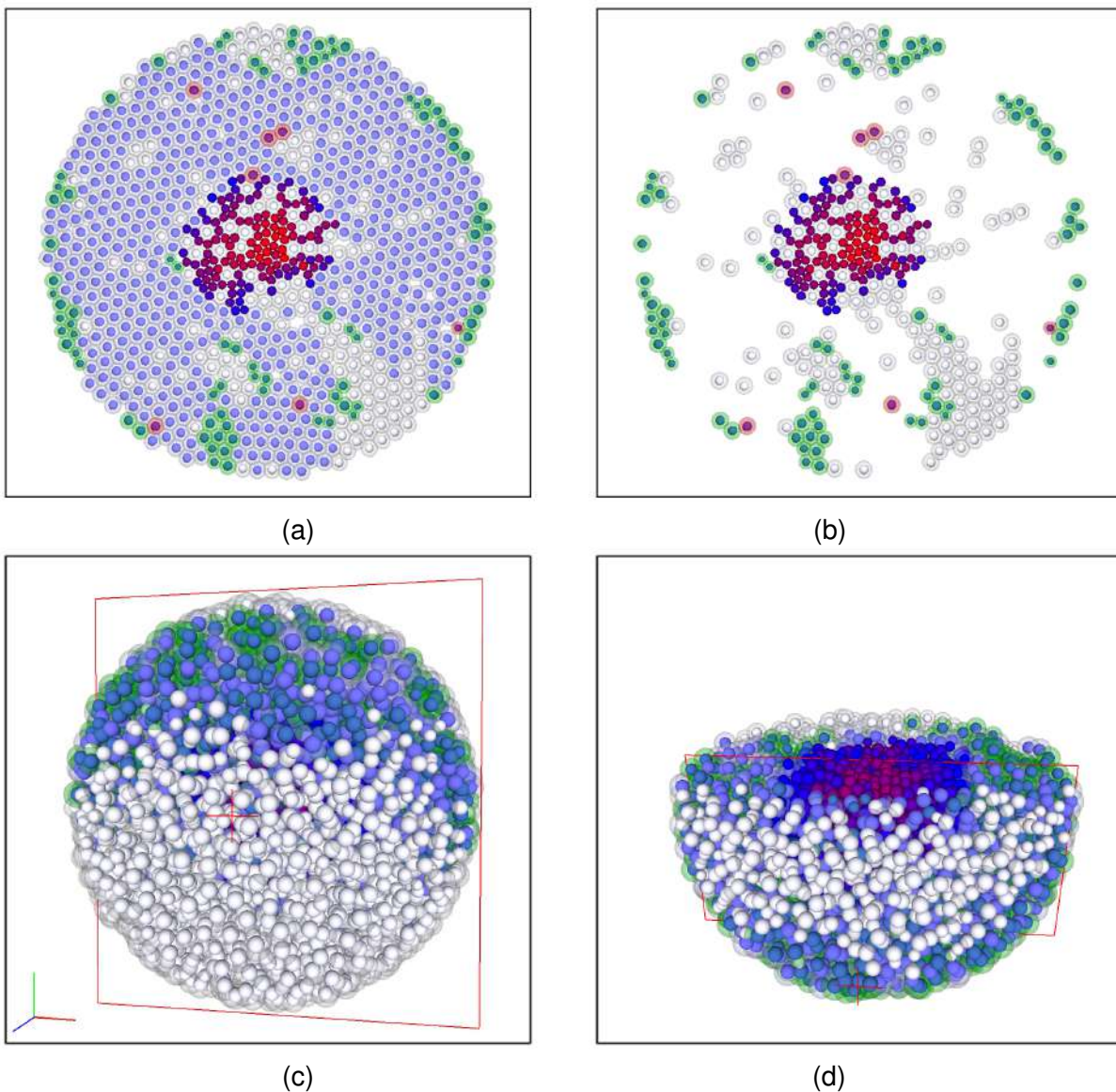


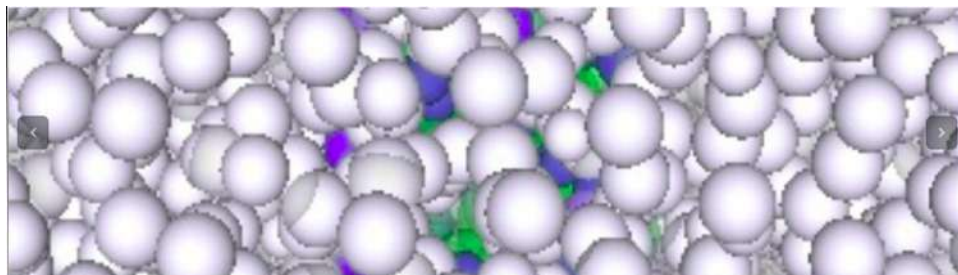
Figura 2: (a)-(b) Visualização bidimensional. (c)-(d) Visualização tridimensional.

na posição inicial da câmera, centralização, velocidade e limitações de distância do *zoom*, que facilitaram a movimentação tridimensional. Também foram implementadas funções de dinamismo, tornando todas essas modificações adaptáveis ao tamanho do modelo no visualizador.

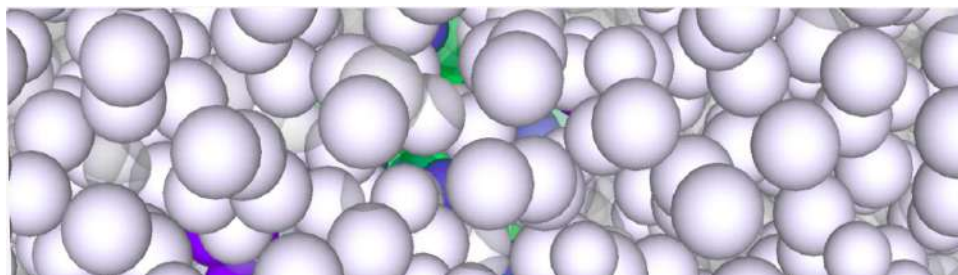
- Com o visualizador sendo facilmente manipulado pelo usuário, foi feito um estudo para se implementar a função de exportação de uma imagem em alta resolução. O visualizador utilizava da memória da tela para exportar a imagem em um arquivo de formato PNG, o que limitava a resolução do arquivo à resolução do monitor em que estava sendo executado. Foi então decidido que um FBO (*Frame buffer object*) seria implementado para aumentar a resolução em que o visualizador era executado. Agora, de acordo com as configurações impostas pelo usuário, a resolução da imagem exportada pode chegar a resoluções bem mais altas, como mostram as Figuras 3(a)-(b).
- Funções para captura de imagem, relacionada a visualização do crescimento do câncer também foram implementadas, pois existiam problemas ao gravar com fluidez esse tipo de visualização. Por isso foram adicionadas funções de captura de imagem em sequência para posteriormente a criação de um vídeo.

4 Conclusões

As implementações realizadas no visualizador tornaram-no mais eficiente e simples de ser manipulado pelo usuário. A qualidade da imagem exportada pode agora ser adequada ao padrão



(a)



(b)

Figura 3: Comparação entre a imagem com a resolução do monitor (1366×768) e depois da implementação do FBO em $8k$.

de publicações científicas. A implementação de função de captura de imagem permite a visualização em detalhes do crescimento do tumor, incluindo a diversidade de componentes que estão presentes no microambiente tumoral.

Referências

- [1] ACS. Atlanta: American Cancer Society. *Cancer Facts & Figures 2016*, 2016.
- [2] A. C. M. de Resende. *Sensitivity Analysis as a Tool for Tumor Growth Modeling*. Master's thesis, Laboratório Nacional de Computação Científica (LNCC/MCTI), Petrópolis/RJ, Brazil, Fevereiro, 2016.
- [3] J. H. L. Fabián. *Calibração de Modelo Matemático do Crescimento Tumoral*. Master's thesis, Laboratório Nacional de Computação Científica (LNCC/MCTI), Petrópolis/RJ, Brazil, Agosto, 2016.
- [4] INCA. Instituto Nacional de Câncer José Alencar Gomes da Silva, Jan. 2016.
- [5] E. A. B. F. Lima. *Phase-Field Models of Tumor Growth with Angiogenesis*. PhD thesis, Laboratório Nacional de Computação Científica (LNCC/MCTI), Petrópolis/RJ, Brazil, Abril, 2014.
- [6] P. Macklin, M. E. Edgerton, A. M. Thompson, and V. Cristini. Patient-calibrated agentbased modelling of ductal carcinoma in situ (dcis): From microscopic measurements to macroscopic predictions of clinical progression. *Journal of Theoretical Biology*, 301:122–140, 2012.
- [7] H. L. Rocha. *Modelagem Híbrida Multiescala para o Crescimento Tumoral*. Master's thesis, Laboratório Nacional de Computação Científica (LNCC/MCTI) (in Portuguese), Petrópolis/RJ, Brazil, Fevereiro, 2016.
- [8] R. V. M. de Souza. *Modelos de crescimento tumoral espacialmente heterogêneos com aplicação de quimioterapia*. Master's thesis, Laboratório Nacional de Computação Científica (LNCC/MCTI) (in Portuguese), Petrópolis/RJ, Brazil, 2013.

Paralelização do Model-R baseada em Dataflow com Apache Spark

Bolsista: Matheus Machado da Rosa Albuquerque^{1,2}
Orientador: Luiz Manoel Rocha Gadelha Júnior²

¹Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (Cefet/RJ)
Petrópolis – RJ – Brasil

²Laboratório Nacional de Computação Científica (LNCC)
Petrópolis – RJ – Brasil

{matheusm, lgadelha}@lncc.br

PIBIC - 2018

1. Objetivos

O presente estudo tem como objetivo analisar e propor uma forma de aprimorar o desempenho computacional do *framework* Model-R. Com o crescimento exponencial dos dados ecológicos, torna-se necessário que ferramentas de modelagem de nichos ecológicos como o Model-R estejam preparadas para atender a esta demanda, sendo capazes de proferir análises em um tempo hábil. Nesta pesquisa, visa-se também a portabilidade do processamento realizado no Model-R para a arquitetura *Big Data* com o intuito de gerar ganho de desempenho mesmo em sistemas de processamento paralelo com arquitetura simplificada.

2. Introdução

O nicho ecológico [Drake et al. 2006] é um conceito biológico fundamental. A modelagem de nichos de espécies é necessária para inúmeras aplicações ecológicas, incluindo a previsão de invasões de espécies, a identificação de reservatórios para doenças, o projeto de reserva natural e a previsão dos efeitos das alterações climáticas antropogênicas e naturais nas faixas das espécies. *Softwares* de modelagem de nicho e visualização de resultados são importantes para identificar e gerar soluções adequadas no quesito da conservação da biodiversidade. A modelagem de nicho pode ser entendida como uma forma de uso e combinação de dados de ocorrência de espécies e variáveis preditivas para inferir projeções derivadas de modelos estatísticos e teóricos. A técnica de modelagem de nicho ecológico em si é vasta e pode ser dividida em diversas etapas. Entretanto, com o avanço dos métodos empenhados na coleta e análise de dados ecológicos, o volume de dados tal como a precisão exigida nos estudos tem crescido gradativamente de forma a não escalar bem com o funcionamento das ferramentas atuais de modelagem de nicho ecológico, exigindo cada vez mais eficiência nas execuções.

O Model-R [Sánchez-Tapia et al. 2018] é um *framework* desenvolvido em linguagem R que nasceu com o objetivo de agrupar ferramentas de modelagem de nichos ecológicos em um único ambiente integrado e amigável onde fosse possível automatizar as principais etapas da análise, como o processo de modelagem e a recuperação dos dados de ocorrência de espécies. A automação e estruturação dos processos do Model-R

influenciam diretamente em seu desempenho, uma vez que estão ligados às fases de processamento e análise de resultados gerados pela modelagem. Outro diferencial é a solução de integrar todas as ferramentas necessárias ao processo de modelagem no próprio *framework*, para que sejam realizadas todas as etapas essenciais, desde a limpeza de dados bem como a modelagem propriamente dita. Soluções como estas acrescentam mais precisão e velocidade à ferramenta tornando-a mais eficiente de modo geral. Atualmente, o Model-R conta com sete algoritmos de modelagem como BIOCLIM, distância Mahalanobis, Maxent, GLM, RandomForest, SVM e DOMAIN. A aplicação também conta com diversas funções para modelagem de nicho ecológico, tal como funções para recuperação de registros de ocorrências de espécies. Devido a sua flexibilidade de uso, Model-R pode ser explorado via sua plataforma *web* bem como em linha de comando.

Na versão mais atualizada da ferramenta foi empregada técnicas de paralelismo tradicionais. Como a aplicação de uma função sobre os elementos de uma lista de forma paralela, com o intuito de melhorar o desempenho. Para isto, foi utilizado o pacote *Snowfall* [Knaus 2010] e foram feitas alterações para que fosse empregado o paralelismo por espécie, ou seja, a análise é particionada de acordo com a quantidade de espécies. Vale destacar também uma outra otimização já implementada, a criação de funções separadas referente a cada uma das etapas de análise, deixando assim, todo código mais modular e abrindo oportunidades de aplicação de paralelismo, tal como o particionamento por algoritmo onde cada *thread* ou nó fica responsável por executar exclusivamente um dos algoritmos da plataforma.

3. Metodologia

Para o desenvolvimento do presente estudo tem-se como estratégia utilizar técnicas de paralelismo computacional através da plataforma Apache Spark [Zaharia et al. 2016] usufruindo de sua API em linguagem R. O Spark nasceu como um mecanismo unificado para processamento de dados distribuídos. A ferramenta possui um modelo de programação semelhante às funções de *Map* e *Reduce*, porém expande esse modelo com uma abstração para compartilhamento de dados chamada *Resilient Distributed Datasets* (RDDs). Para o desenvolvimento do código foi proposto o uso dos pacotes *SparkR* [Venkataraman et al. 2016] e *Sparklyr*¹, que fornecem suporte necessário para o uso e integração do Spark a linguagem R. Mesmo com o desempenho satisfatório, é necessário que o Model-R esteja preparado para o futuro devido a crescente demanda por maiores velocidades de processamento e ao constante aumento na quantidade dos dados ecológicos. Outra questão que deve ser levada em consideração é a portabilidade da ferramenta para a arquitetura *Big Data*, visando escalabilidade em sistemas de processamento mais simples que *clusters* tradicionais para computação de alto desempenho (HPC).

Esse tipo de arquitetura é conhecida como *Shared-Nothing*. Em geral, *clusters* para HPC usufruem de um sistema de arquivos compartilhado onde todos os nós conseguem acessá-lo. Este tipo de arquitetura, por sua vez, dispõe de nós com suas próprias unidades de armazenamento. Para processamento nesse tipo de plataforma, os dados são distribuídos, processados e por fim agrupados para uma análise final. A vantagem desse tipo de arquitetura é o custo dos equipamentos, uma vez que *clusters* tradicionais necessitam de sistemas e *hardware* mais sofisticado, enquanto *clusters Big Data* podem ser en-

¹<https://cran.r-project.org/web/packages/sparklyr/index.html>

xergados como um conjunto máquinas comuns processando dados locais paralelamente. A maioria das aplicações do tipo *Big Data* necessitaria combinar vários tipos de processamentos diferentes, isso se deve à própria natureza desse tipo de dado que é diversificada e não-estruturada. Desta forma, modelos como este poderiam introduzir complexidade e gerar ineficiência por parte dos mecanismos de processamento especializado. Ferramentas como o Spark e arquiteturas típicas para processamento de dados *Big Data* também se diferenciam em diversos aspectos em relação aos tradicionais.

Para a implementação da estrutura para modelagem de nichos presente no Model-R de forma paralelizada foram levados em consideração algumas questões como: a natureza dos dados e também o plano de paralelismo a ser empregado. A principal limitação encontrada durante o estudo foi a questão da própria natureza diversificada dos dados necessários à modelagem, que variam de arquivos de variáveis preditivas em formato *Raster Stack*² até arquivos de ocorrência em formato CSV. Para isto, foram criados diversos *Data Frames* de acordo com a necessidade de cada etapa de processamento, convertendo e encapsulando os dados necessários a análise no próprio *Data Frame*. Os *Data Frame* podem ser compreendidos como uma coleção de dados distribuídos e organizados em formato de tabular, sendo possível armazenar dados de diversas naturezas. Em virtude desse formato de dado ser nativo ao Spark, torna-se possível realizar toda manipulação e análise necessária via Spark, ganhando em performance. Foi possível desenvolver versões otimizadas dos algoritmos BIOCLIM e SVM aplicando técnicas de paralelismo computacional nas etapas de processamento de cada algoritmo.

3.1. BIOCLIM

O BIOCLIM é um algoritmo clássico de modelo de envelope climático. Seu funcionamento se apoia no cálculo da similaridade de uma localização, comparando os valores das variáveis ambientais em qualquer localização com uma distribuição percentual dos valores em locais conhecidos de ocorrência. Em especial, para o desenvolvimento dessa versão do algoritmo foram utilizadas apenas consultas SQL via SparkR. Foi necessário criar dois *Data Frames*, um referente ao arquivo de variáveis preditivas e outro, referente ao arquivo de ocorrências. Como o arquivo original de variáveis preditivas se encontra em formato *Raster Stack*, ou seja, um arquivo com várias matrizes (*layers*), foi necessário convertê-lo para um formato tabular para realização de consultas. Com isso, foi realizada a concatenação vertical de todas as matrizes, resultando em uma única matriz.

Uma das etapas do BIOCLIM é mapear esses pontos geográficos (latitude, longitude) presentes no arquivo de ocorrências nos valores das variáveis ambientais (linha, coluna) das matrizes. A primeira consulta tem esse objetivo e retorna um *Data Frame* contendo uma coluna com o nome das espécies e outras duas com os valores de linha e coluna, referente aos índices das matrizes. A segunda consulta consome esse *Data Frame* gerado e juntamente com os dados de ocorrência gera um novo *Data Frame*, contendo o nome das espécies e o valor das variáveis preditivas em todos os *layers*. Por fim, é gerado o envelope contendo os valores de máximo e mínimo das variáveis ambientais para cada espécie.

²<https://www.rdocumentation.org/packages/raster/versions/2.6-7/topics/stack>

3.2. SVM

O *Supporting Vector Machine* (SVM) é um método de aprendizagem de máquina capaz de analisar e reconhecer padrões em dados e, nesse caso, foi usado para a classificação. Para desenvolver essa versão do algoritmo foi utilizada a versão paralelizada do SVM disponível no Spark, derivada da biblioteca MLlib [Meng et al. 2016]. Inicialmente é criado um *Data Frame* contendo todos os dados de ocorrência das espécies, em seguida, é acrescentada uma coluna referente a presença da espécie no determinado ponto. Essa coluna foi completamente preenchida com “Yes”. Depois é gerado um segundo *Data Frame*, de tamanho ajustável, contendo pontos de ocorrência gerados aleatoriamente e com a coluna de presença completamente preenchida com “No”. Esse segundo *Data Frame* é essencial à classificação devido ao SVM linear necessitar de dados de ausência também. Por fim, esse *Data Frame* é concatenado verticalmente ao primeiro gerando uma matriz única.

Para executar o SVM, uma parte dos dados é separada para treino e o restante para teste. O modelo é criado usando os dados de treino e a classificação é feita utilizando a coluna de presença em relação a cada espécie. A fim de testar a precisão da classificação, é realizada a predição do modelo com os dados de teste. Finalmente é criada a matriz de confusão com base no resultado da predição em relação à coluna presença dos dados de teste, para visualizar o número total de acertos, falsos positivos e falsos negativos. Visando uma análise mais detalhada, foram desenvolvidas 3 consultas SQL para visualizar individualmente o número de acertos, falsos positivos e falsos negativos de cada espécie.

4. Resultados e Discussão

A capacidade de desempenho e escalabilidade do Apache Spark é promissora e pode ser comprovada durante o estudo. Os testes de desempenho iniciais foram realizados sobre as consultas SQL presentes na versão do BIOCLIM desenvolvido. Em um primeiro momento, as consultas foram executadas com suporte no Spark e em seguida de forma procedural utilizando a biblioteca *sqldf*³. O *cluster* Buriti utilizado durante a análise, está localizado no Laboratório Nacional de Computação Científica e conta com 23 processadores AMD Opteron(TM) 6238 de 2.6GHz com 12 cores e também 75 Gigabytes de memória RAM.

A primeira consulta, que tem por objetivo mapear os valores das coordenadas em índices, consumiu um *Data Frame* de 69 Kilobytes referente ao arquivo de ocorrências e levou 77 milissegundos para ser carregado em uma RDD. A consulta com suporte no Spark levou 2 segundos para ser totalmente processada enquanto a de modo procedural levou 0,09 segundos. A segunda consulta, referente a criação do *Data Frame* contendo o nome das espécies e os valores das variáveis ambientais, consumiu o *Data Frame* de 16 Kilobytes resultante da primeira consulta SQL e também o *Data Frame* de 16 Megabytes referente ao arquivo de variáveis preditoras, que levou 0,9 segundos para ser carregado em uma RDD. Esta segunda consulta levou 3 minutos para ser totalmente processada enquanto a de modo procedural levou 53,43 segundos.

O baixo desempenho nas consultas com suporte no Spark pode ser justificado devido ao tamanho pequeno dos arquivos consumidos. Ferramentas como esta apresentam

³<https://cran.r-project.org/web/packages/sqldf/README.html>

grande vantagem computacional quando processam grandes quantidades de dados, isto ocorre devido ao tempo necessário para a distribuição e escrita dos dados em todas as unidades de armazenamento de cada nó presente no *cluster*. Este fato pôde ser comprovado nos testes realizados pela equipe de desenvolvimento do Spark, onde foi aplicada a regressão logística a 100 Gigabytes de dados. O Spark foi capaz de processar o cálculo em alguns segundos enquanto o Hadoop levou cerca de 2500 segundos.

O teste com SVM tinha por objetivo medir o desempenho como também a precisão da classificação realizada. O estudo envolveu o arquivo de ocorrências com 56 Kilobytes. Primeiramente foi criado um *Data Frame* de 61 pontos de ausência gerados aleatoriamente. O tamanho desse arquivo é referente a 10% do tamanho do arquivo de ocorrências. Após concatenar os dois *Data Frames*, 20% foi separado para treino e o restante para teste. Das 544 ocorrências de teste, o modelo classificou corretamente no total 500, gerou 44 falsos positivos e não apresentou falsos negativos. Isso significa um percentual de erro de apenas 8,08%. Dos quatro tipos de espécies presentes no arquivo de ocorrência, o modelo conseguiu classificar corretamente todas as ocorrências referentes a uma das espécies e as demais com taxas de erro inferiores a 17%.

5. Conclusões

Com isso, avalia-se a possibilidade de implementação de técnicas de paralelismo às etapas de modelagem do Model-R. Assim estima-se tornar a ferramenta computacionalmente superior em desempenho de forma a escalar melhor com conjunto de dados maiores e também proporcionar a portabilidade para arquiteturas do tipo *Big Data*, não necessitando de sistemas de processamento paralelo com arquiteturas mais sofisticadas. Como trabalho futuro visa-se a continuidade do estudo de forma a criar versões otimizadas para todos os algoritmos presentes no Model-R.

Referências

- Drake, J. M., Randin, C., and Guisan, A. (2006). Modelling ecological niches with support vector machines. *Journal of applied ecology*, 43(3):424–432.
- Knaus, J. (2010). Developing parallel programs using snowfall. Retrieved from CRAN.
- Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkataraman, S., Liu, D., Freeman, J., Tsai, D., Amde, M., Owen, S., et al. (2016). Mllib: Machine learning in apache spark. *The Journal of Machine Learning Research*, 17(1):1235–1241.
- Sánchez-Tapia, A., de Siqueira, M. F., Lima, R. O., Barros, F. S. M., Gall, G. M., Gadelha, L. M. R., da Silva, L. A. E., and Osthoff, C. (2018). Model-r: A framework for scalable and reproducible ecological niche modeling. In Mocskos, E. and Nesmachnow, S., editors, *High Performance Computing*, pages 218–232, Cham. Springer International Publishing.
- Venkataraman, S., Yang, Z., Liu, D., Liang, E., Falaki, H., Meng, X., Xin, R., Ghodsi, A., Franklin, M., Stoica, I., and Zaharia, M. (2016). Sparkr: Scaling r programs with spark. In *Proceedings of the 2016 International Conference on Management of Data, SIGMOD '16*, pages 1099–1104, New York, NY, USA. ACM.
- Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., Meng, X., Rosen, J., Venkataraman, S., Franklin, M. J., et al. (2016). Apache spark: a unified engine for big data processing. *Communications of the ACM*, 59(11):56–65.

1 - Dados Gerais:

Título: Controle ótimo de vigas de Euler - Bernoulli

Bolsista: Monique Ribeiro da Costa.

Orientador: Jaime Edilberto Muñoz Rivera.

Bolsista de Iniciação Científica- CNPq.

Período do desenvolvimento: 02/02/2018 à 25/07/2018

2 - Objetivos:

- Formular o problema de controle ótimo de uma viga no modelo de Euler Bernoulli, onde minimiza-se um funcional quadrático que otimiza a deformação de acordo com uma função objetivo;
- Proporcionar contato com a ferramenta de programação MALTLAB;
- Proporcionar contato com técnicas matemáticas mais elaboradas.

3 - Introdução:

Sabendo que muitas estruturas são construídas usando vigas, as quais defletem ou distorcem por decorrência do próprio peso ou de uma força externa, podendo ocasionar ruptura dependendo da intensidade da força exercida ou pela constituição de seu material. Simularemos las deformaciones y el cuadro de las tensiones de la viga de comprimento L usando o modelo de Euler Bernoulli. Este modelo é baseado no seguinte princípio:

“A posição de equilíbrio dos corpos elásticos é aquela que minimiza sua energia potencial elástica.”

Usando este princípio obtemos que a posição de equilíbrio da viga definida pelo gráfico da função u que é solução da seguinte equação:

$$\alpha u + \beta u_{xxxx} = f \quad (1)$$

Verificando a condição de contorno:

$$u(0) = u_x(0) = 0, u(L) = u_x(L) = 0 \quad (2)$$

Sendo f a força distribuída que está atuando sobre a viga. Assim, a função u será aquela cujo gráfico define a posição de equilíbrio da viga. Então para cada f temos uma posição de equilíbrio u .

O problema de controle de vigas consiste em controlar a solução u da equação descrita em (1). Deve-se escolher a força f de tal forma que a posição de equilíbrio esteja mais parecida com a função objetivo Z_d . Isto é, como deve ser f de tal forma que $u(x) - z_d(x)$ seja o menor possível. Aqui aparece um problema de minimização sobre o conjunto de todas as possíveis forças distribuídas que podem agir sobre a viga.

4 - Material e Métodos ou Metodologia:

Utilizamos os métodos dos espaços de Hilbert e de minimização de funcionais quadráticos sobre estes espaços. Estudamos o problema de controle ótimo da equação (1). O problema consiste em determinar a força que deve ser concentrada acima da viga de tal forma que a viga esteja próxima da nossa função objetivo. O funcional custo será dado por um funcional quadrático e as forças atuantes sobre a viga serão os controles.

O problema tende a encontrar o controle ótimo que minimiza o funcional custo, e a caracteriza-lo de tal forma que seja possível calcular numericamente, para isto precisa-se encontrar o sistema de otimalidade que consiste na equação de estado e a equação do sistema adjunto.

A solução numérica do problema foi desenvolvida usando o método dos elementos finitos baseados nas diferenças centrais, método a qual aproxima-se a derivada de uma função pela sua diferença.

5 - Resultados e Discussão:

Caracterizamos o controle ótimo através do sistema de otimalidade que consiste na equação de estado e a equação do sistema adjunto, calculamos a solução numericamente, utilizando o método de diferenças finitas baseadas nas diferenças centrais, resolvimos assim o sistema discreto computacionalmente, utilizando o método de eliminação de Gauss, em um programa feito em MATLAB®, obtendo desta forma os valores numéricos da deformação da viga em cada ponto da discretização dela. Utilizando a plataforma gráfica do MATLAB® para simular as deformações

6 - Conclusões:

- 1) Deduzimos o modelo da viga de Euler Bernoulli usando argumentos físicos da teoria de elasticidade;
- 2) Calculamos soluções aproximadas do modelo de Euler Bernoulli usando diferenças finitas;
- 3) Definimos o problema de controle ótimo para vigas unidimensionais onde tentamos controlar através de um funcional quadrado a deformação de viga;
- 4) Concluimos que para funcionais devidamente “coercivos” o problema de controle ótimo possui existência e unicidade;
- 5) O problema numérico resultado da caracterização do problema de controle da viga de Euler Bernoulli é bem complexo devido o sistema ser de quarta ordem, dificultando muito as aproximações numéricas, criando matrizes de grande porte quando consideramos mais nós no intervalo. Estamos estudando a possibilidade de substituir o modelo de Euler-Bernoulli pelo modelo de vigas de Timoshenko.

7 - Referências bibliográficas:

RIVERA, JAIME . Título:Cálculo Difrencial II e equações diferenciais.

RIVERA, Jaime Edilberto Muñoz. Métodos Computacionales Aplicados a las Ecuaciones Difrencales.

Editora Thomsom.

ZILL, DENNIS G. Equações Diferenciais com Aplicações em Modelagem. XXX: Thomson, 2003.

SOUZA, Salete. Disciplina: Resistência dos materiais. Escola de engenharia metalúrgica de Volta Redonda. Universidade Federal Fluminense sd. Disponível em <<http://www.professores.uff.br/salete/res1/aula10.pdf>

Avaliação de Métodos de Aprendizado em Aplicações de Saúde

Nome do Bolsista: Raquel de Abreu Junqueira Gritz

Nome do Orientador: Fábio André Machado Porto

Tipo de Bolsa e Período do Relatório: Bolsa de Iniciação Científica – IC / PIBIC – Julho de 2018

I – Objetivos

Este trabalho de iniciação científica tem como objetivo comparar as diferentes técnicas de aprendizado aplicadas a problemas ligados à predição, em particular considerando dados de saúde, no contexto da cooperação LNCC-DEXL / FIOCRUZ-ICICT, e identificar características dos métodos que melhor se adéquem ao problema.

II – Introdução

A busca pela capacidade de analisar um grande volume de dados por meio de métodos estatísticos e o uso de uma variedade de algoritmos para encontrar padrões nos dados, mostra que, com base nesses padrões, é possível realizar predições.

Uma área em que predições são bem vindas é a saúde. Atualmente, um grande número de dados encontra-se disponível e que pode ser usado em um processo de predição.

Neste contexto, este trabalho, em uma primeira fase, utilizou dados associados à mortalidade infantil no subconjunto dos municípios brasileiros. Em particular utilizou-se como alvo do modelo de predição o atributo que identifica o número de crianças indo a óbito em até 5 anos.

Iniciamos o projeto estudando a técnica de aprendizado supervisionado baseado em árvore de decisão. Como processo de aprendizado, a técnica requer uma passagem pela etapa de treinamento, onde um modelo na forma de árvore de decisão é construído, seguido por um processo de predição em que um dado não visto no processo de treinamento é fornecido como entrada para o processo de predição com o modelo criado. O resultado espelha-se em um dos grupos de classificação que mais se coincide com os valores do dado de entrada.

Na continuação deste trabalho, pretendemos avaliar a técnica de árvores de decisão sobre os dados de saúde do DATASUS sobre Hospitalizações, com o objetivo de predizer o estado final do paciente quando der entrada em uma internação.

III – Material e Métodos ou Metodologia

Esse trabalho considerou dados de municípios relacionados à mortalidade infantil, tendo como tarefas:

- a) aprendizado da linguagem Python com o foco voltado para as bibliotecas Pandas e Scikit-learn de aprendizado de máquina;
- b) obtenção dos dados e identificação das fatias de treinamento e teste;

- c) estudo das técnicas de aprendizado e dos algoritmos disponíveis para Python (scikit-learn), considerando as árvores de decisão;
- d) preparo dos dados segundo as técnicas empregadas;
- e) realização de treinamento e teste para as técnicas selecionadas;
- f) apresentação do estado de desenvolvimento do trabalho para o grupo DEXL Lab.

IV – Resultados e Discussão

Neste trabalho foi usado o algoritmo de Árvore de Decisão em Python com Scikit-learning tendo como base dados relacionados a saúde de 5565 municípios brasileiros.

Teve como objetivo retornar uma árvore de boa precisão que determinaria as causas de uma alta taxa de mortalidade infantil até os 5 anos.

Uma função foi criada para determinar os valores da mortalidade infantil organizados em faixas de níveis alto, médio e baixo.

O Dataset de 46 colunas foi dividido em 42 colunas para treinamento e 1 para teste e foi possível retornar a matriz de confusão que teve a precisão aproximada de 0.99 para valores considerados altos, 0.80 para valores médios e 0.95 para valores baixos.

Como se pode ver abaixo, dos 384 valores determinados como altos, o algoritmo identificou 361 de forma correta, dos 371 valores médios, ele retornou 354 corretamente, enquanto que dos 358 de valores baixos, ele acertou 291.

```
[[361 0 23]
 [ 0 291 67]
 [ 3 14 354]]
```

Valor Predito	alto	baixo	medio	All
Valor Real				
alto	361	0	23	384
baixo	0	291	67	358
medio	3	14	354	371
All	364	305	444	1113

Foi possível assim chegar ao que era pretendido inicialmente, compreendendo a importância da identificação dos dados a serem separados e tendo como resultado uma boa precisão.

V – Conclusão

Nesta fase inicial foi possível criar um modelo de aprendizado baseado na técnica de Árvore de Decisão para classificação de atributos ligados à Mortalidade Infantil capaz de classificar novas entradas como com características de mortalidade: alta, média ou baixa.

Temos como objetivo futuro, aprimorar os estudos e fazer a análise de dados para saúde onde se buscará fazer a predição relacionada a dados de Hospitalização, determinando características que possam levar a alta de um paciente ou ao óbito do mesmo.

VI – Referências Bibliográficas

Thomas G. Dietterich, Ensemble Methods in Machine Learning, ISBN: 3-540-67704-6, Springer-Verlag London, UK, 2000

Bolsista:

René Constancio Nunes de Lima

Orientadores:

Elson M. Toledo(LNCC)

Regina Célia P. Leal Toledo (UFF- co-orientadora)

Relatório parcial do projeto:

**SIMULAÇÃO NUMÉRICA E COMPUTACIONAL DO TRÁFEGO
VIÁRIO**

LNCC

PROGRAMA PIBIC/PIBITI

Período

Abril a Julho de 2018

1- Objetivos

O objetivo do projeto que vem sendo desenvolvido consiste em analisar o tráfego viário utilizando modelos microscópicos, baseados em Autômatos Celulares (AC), considerando, em todas as análises, a influência de diversas políticas de aceleração e desaceleração, que é o que diferencia o presente trabalho.

Em uma das vertentes do projeto, o tradicional modelo probabilístico de AC proposto por Nagel e Schreckenberg [1], chamado modelo NaSch, foi modificado para possibilitar a consideração de diversas políticas ao dirigir, com a inclusão de diferentes possibilidades de valores de aceleração/desaceleração a cada instante de tempo. Na análise aqui considerada, essa modificação consiste em utilizar uma discretização mais refinada do que a proposta usualmente para o modelo NaSch e uma Função de Densidade de Probabilidade (FDP) não uniforme. A FDP Beta foi utilizada para definir essas diferentes políticas de aceleração e desaceleração. Essas políticas variam de uma aceleração mais cautelosa, que tem a tendência de acelerar/desacelerar mais suavemente, a uma mais agressiva, que acelera e desacelera mais bruscamente, e são dadas pelas diferentes médias da FDP utilizada. Resultados dessa abordagem, com a participação de orientados PIBIC/LNCC, têm sido apresentados em congressos e periódicos [2-4].

O presente relatório, que descreve as atividades iniciadas em Abril do presente ano, tem o objetivo de avaliar a influência de diferentes discretizações da via nos resultados obtidos, bem como analisar a influência da variância da FDP na definição dessas políticas.

2- Introdução

Minimizar os grandes engarrafamentos, que aumentam a cada dia nos grandes centros e arredores, é um dos fatores importantes nos dias de hoje quando se pensa em qualidade de vida, redução do *stress* e da poluição sonora e do ar. Embora um congestionamento possa ser causado por um acidente, um gargalo ou motoristas simplesmente mudando de faixas em estradas movimentadas, muitas das vezes ele tem origem na forma como os motoristas reagem nas diversas situações do tráfego.

Pesquisadores da área, como Dr. J. Laval [5], que conduz uma pesquisa sobre motoristas agressivos e tímidos, afirma que os motoristas agressivos, que dirigem muito rápido e muito perto do veículo à frente, ou motoristas tímidos, que deixam uma lacuna muito grande entre veículos, enviam uma "onda de desaceleração" para trás, podendo levar a grandes engarrafamentos que impedem o tráfego. Afirma ainda que a característica "*stop and go*", além de causar incômodo e *stress* aos motoristas, também aumenta o consumo de combustível e as emissões de gás, e impõe riscos à segurança nas vias [6]. No entanto, os modelos propostos na literatura têm tratado essa questão apenas limitando a velocidade máxima do perfil ou, no máximo, tratando de uma forma diferenciada a distância entre veículos.

Nesse trabalho, apresentamos um estudo que vem sendo desenvolvido propondo uma modificação no modelo NaSch que possibilite a representação de diferentes valores de tendências de aceleração, dependendo do perfil do motorista. Para isso, foi utilizada uma discretização mais refinada do que a proposta para o modelo NaSch tradicional, e uma Função de Densidade de Probabilidade (FDP) não uniforme, a FDP Beta, para descrever o quanto cada motorista gostaria de acelerar/desacelerar. Tal função possui parâmetros de entrada que geram resultados com diferentes médias, que definem esses diferentes perfis de aceleração.

A influência desses diferentes perfis na dinâmica do fluxo viário, bem como na existência de situações perigosas na via que podem conduzir a acidentes são resultados já obtidos pelo grupo de pesquisa [2-4]. No entanto, a importância da discretização da via e das diferentes FDP's Beta com a mesma média, e diferentes variâncias, é um trabalho ainda não realizado, que é no que estamos trabalhando no momento.

3- Material e Métodos ou Metodologia

3.1 Metodologia

Iniciei minha participação no grupo de pesquisa em abril do presente ano, com projeto já em andamento. Então, nessa fase inicial, precisei ter conhecimento do problema considerado e da técnica utilizada, e definir o tipo de trabalho a ser inicialmente desenvolvido para colaborar com os novos resultados do projeto. Para isso foi utilizada a seguinte metodologia:

- (1) Inicialmente comecei estudando a bibliografia recomendada, entre elas a tese de Doutorado de M.P. Zamith [7], que faz, em seus capítulos iniciais, uma excelente descrição do problema e também da modelagem microscópica, particularmente a baseada em Autômatos Celulares. Também foram lidos os trabalhos que vêm sendo desenvolvidos pelo grupo de pesquisa [2,4], particularmente aqueles que têm participação de bolsistas de Iniciação Científica;
- (2) Para ter conhecimento do funcionamento da técnica utilizada, rodei o programa desenvolvido pelo grupo de pesquisa, com o algoritmo do NaSch original, variando alguns parâmetros de entrada para testar seu funcionamento, variando condições iniciais, tamanho da via e diferentes probabilidades de aceleração/desaceleração dadas pelo modelo original;
- (3) A seguir, avaliamos qual contribuição poderia ser dada por alguém que está iniciando sua participação e decidimos que isso se daria em duas vertentes:
 - (a) Um estudo para avaliar a influência da discretização da via nos resultados alcançados, que é uma análise não usual nessa técnica;
 - (b) Avaliar a influência que as diferentes variâncias das FDP's utilizadas têm na definição dessas políticas de aceleração. Essas diferentes políticas são dadas por diferentes curvas da FDP Beta, com diferentes médias, diretamente relacionadas à discretização utilizada e ao perfil adotado.

3.2 Modelo

(i) O modelo NaSch:

O modelo NaSch [1] foi o primeiro e mais conhecido modelo probabilístico de AC. Nesse modelo, em sua proposta tradicional, a via é unidimensional, discretizada em células, onde cada célula ou é ocupada por um veículo ou está vazia. Cada veículo ocupa uma célula, normalmente considerada com $7,5m$, e o tempo também é discretizado, sendo cada instante de tempo t considerado 1 segundo. Cada veículo ocupa a posição i , no instante de tempo t , ou seja, posição $x(i, t)$ e velocidade dada por $v(i, t)$. A distância entre um veículo e outro, dada pelo número de células vazias na frente de cada veículo, é definida como $d(i, t) = x(i+1, t) - x(i, t) - L(i)$, sendo $L(i)$ o número de células ocupadas por um veículo i , que no modelo tradicional, $L(i) = 1$ para todos os veículos i . O algoritmo consiste em um conjunto de quatro regras, aplicadas paralelamente a todos os veículos, a cada segundo.

Algoritmo 1

-
- (1) *Aceleração:* $v(i, t+1) = \min[v(i, t) + A, V_{max}]$
 - (2) *Desaceleração:* $v(i, t+1) = \min[v(i, t+1), d(i, t)]$
 - (3) *Desaceleração:* $v(i, t+1) = \max[v(i, t+1) - A, 0]$, com uma probabilidade p .
 - (4) *Movimento:* $x(i, t+1) = x(i, t) + v(i, t+1)$
-

sendo V_{max} a velocidade máxima da via, $A = 1$ a aceleração dos veículos e p , que modela a incerteza no comportamento dos motoristas, a probabilidade do motorista não querer acelerar, ou desacelerar, dependendo da situação. Assim, o modelo NaSch consegue simular a ocorrência espontânea de congestionamentos, a relação entre fluxo e densidade e as zonas de fluxo livre,

congestionado mas não representa bem a transição entre essas, definida como região de metaestabilidade. Apesar disso, ele é utilizado, com frequência, para avaliar as diversas situações e diferentes topologias das vias.

(ii) O modelo NaSch modificado:

No modelo NaSch original, $A = 1$ e cada veículo acelera ou desacelera 1 cel/s^2 , o que corresponde a $7,5 \text{ m/s}^2$, que é um valor alto para dados reais. Para gerar diferentes políticas de aceleração, dividimos a célula de $7,5 \text{ m}$ do NaSch em células menores, para que seja possível que cada motorista acelere, caso permitido, de diferentes formas. Assim, o valor da variável A , no Algoritmo 1, é probabilístico e é definido, para cada motorista, pela FDP Beta, com diferentes médias. Essa FDP gera um valor de α ($0 < \alpha < 1$), que é utilizada para calcular a nova aceleração $A = \text{int} [(1-\alpha) A_{max}]$, onde int é uma função que retorna o inteiro mais próximo. Assim, o parâmetro p designa se o motorista vai acelerar ou não, e A como ele irá acelerar: mais suavemente ou mais bruscamente. A FDP Beta utilizada é definida como :

$$\beta(a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1}$$

onde $0 \leq x \leq 1$ e $\Gamma(n+1) = n!$, n um inteiro positivo. Dados os parâmetros a e b , resulta uma função $\beta(a,b)$. Para a geração desses valores utiliza-se a Técnica de Rejeição do Método de Monte Carlo. O valor da coordenada x é o valor de α . Valores de α pertos de 0 produzem acelerações mais próximas de A_{max} , enquanto valores mais próximos a 1 produzem acelerações próximas de 0 . Para cada par de valor a e b , podemos calcular a média $M = \frac{a}{a+b}$. Então, podemos relacionar cada perfil de aceleração a um par de parâmetros a e b , onde o valor médio determina a tendência de aceleração de um perfil. Para exemplificar, se a via for discretizada em células de $1,5 \text{ m}$, é possível modelar o motorista que tende a acelerar/desacelerar mais suavemente, com média de $A = 1$, até o que tende a acelerar/desacelerar mais bruscamente, com média de $A = 4$. E quando A é sempre igual a 5 temos o modelo NaSch, com um veículo ocupando 5 células.

4- Resultados e discussão

Resultados para as principais grandezas (fluxo, densidade, velocidade média) já foram apresentados em trabalhos anteriores, mostrando a importância de se considerar as diferentes políticas de aceleração. Para exemplificar, apresentamos na Figura 1 resultados do modelo NaSch modificado quando dividimos a célula do modelo NaSch ($C = 7,5 \text{ m}$) em 5 células. Assim, 1 célula tem $1,5 \text{ m}$. Cada veículo ocupa 5 células e a velocidade máxima da pista é $V_{max} = 25 \text{ cels/s} = 135 \text{ km/h}$, para poder comparar com resultados do modelo NaSch tradicional. $A_{max} = 5 \text{ cels/s}^2$ e a nova aceleração por unidade de tempo pode obter valores de 0 a 5 . Utilizamos $p = 30\%$, como usual. Para essa discretização, definimos as seguintes políticas de aceleração/desaceleração:

- (1) Agressiva : $\beta(10,30)$ – Acel. média/s = 4
- (2) Intermediária I: $\beta(20,28)$ – Acel. média/s = 3
- (3) Intermediária II: $\beta(28,20)$ – Acel. média/s = 2
- (4) Cautelosa: $\beta(30,10)$ – Acel. média/s = 1

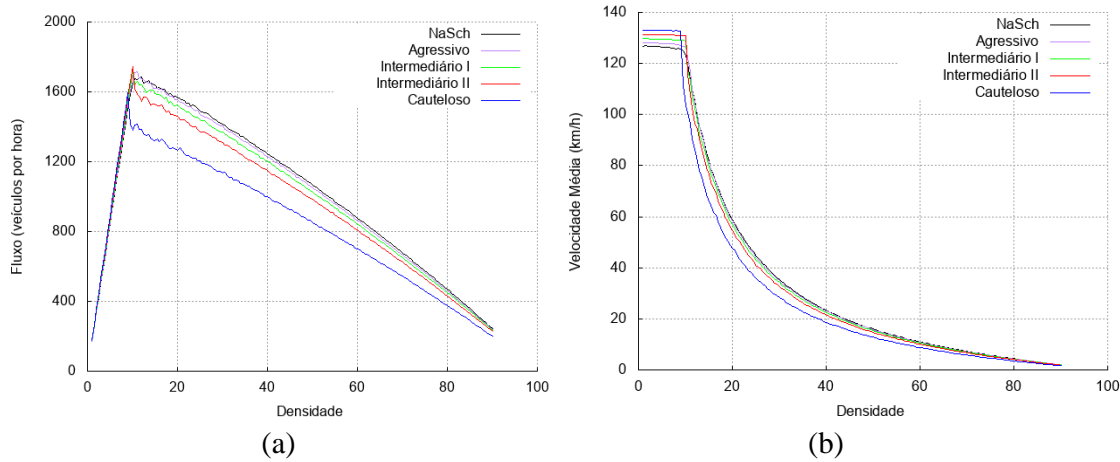


Figura 1: Diagramas (a) fluxo-densidade; (b) velocidade-densidade [2]

Esses resultados, e os outros desse relatório, foram obtidos para uma pista circular, de 15km, durante 15000 segundos. São feitas simulações para densidades ρ de 1% até 90%, com incrementos de 0,5, sendo ρ a porcentagem de células ocupadas da pista. Para cada densidade, são distribuídos veículos aleatoriamente pela pista, de acordo com a densidade, com velocidade $v(i,0)=0$. Para eliminar efeitos transientes, são descartados os 10000 primeiros segundos e considerados os próximos 5000 segundos.

4.1 Mantendo a média e alterando a variância

Para analisar a influência da variância da FDP Beta, na representação das políticas de aceleração, foram comparados resultados com a mesma média e diferentes variâncias, para cada perfil definido pela discretização com células de $1,5 m = C / 5$, onde a variância para a FDP Beta, com parâmetros a e b , é dada por $Var = ab / [(a + b + 1)(a + b)^2]$

Para exemplificar apresentamos, na Figura 2, o diagrama fundamental (fluxo-densidade) para as políticas de aceleração:

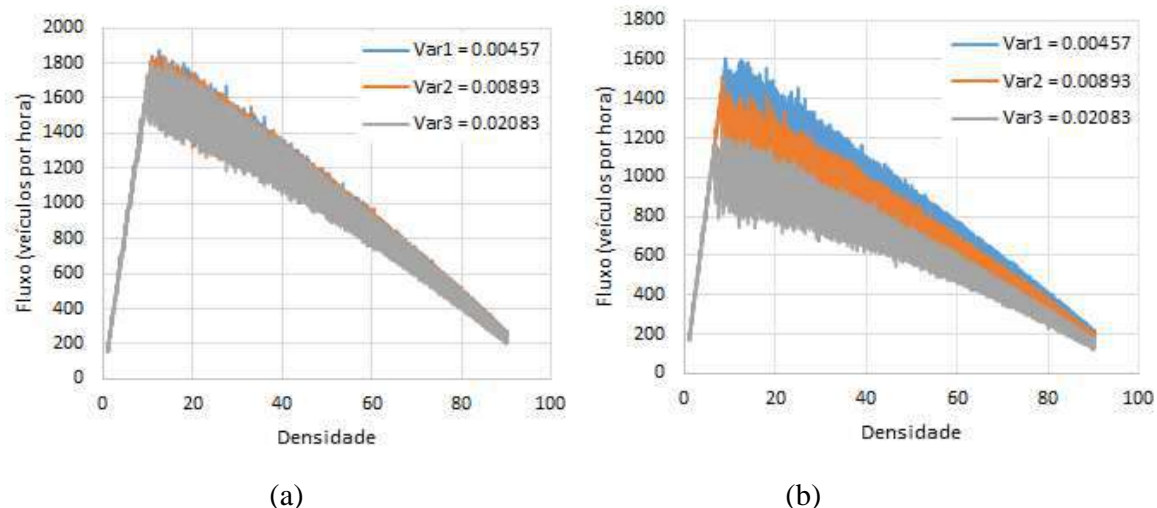


Figura 2: Diagramas fluxo-densidade – influência da variância para política de aceleração (a) agressiva; (b) cautelosa

(i) Agressiva:

$$B(10,30) \rightarrow \text{Var1} = 0.00457; B(5,15) \rightarrow \text{Var2} = 0.00893; B(2,6) \rightarrow \text{Var3} = 0.02083.$$

(ii) Cautelosa:

$$B(30,10) \rightarrow \text{Var1} = 0.00457; B(15,5) \rightarrow \text{Var2} = 0.00893; B(6,2) \rightarrow \text{Var3} = 0.02083.$$

Pelos resultados obtidos, nos gráficos apresentados e em outros obtidos na pesquisa, observa-se que a variância influencia mais, quanto menor for a média da aceleração a cada intervalo de tempo. Com isso, a maior influência se dá para políticas mais cautelosas de aceleração/desaceleração. Nesse caso, continuamos a avaliar a influência da PDF nesses resultados e a melhor forma para considerá-los nas situações já analisadas anteriormente, como na probabilidade de situações perigosas que podem conduzir a acidentes nas vias.

4.2 A influência da discretização da via

Para analisar a influência de diferentes discretizações da via nos resultados consideramos, como referência, o tamanho da célula do tradicional modelo NaSch ($C = 7,5m$). Foram realizadas simulações para discretizações de $C/5$, $C/10$ e $C/15$ e $V_{max} = 135km/h$ para todas as discretizações (Figura 3).

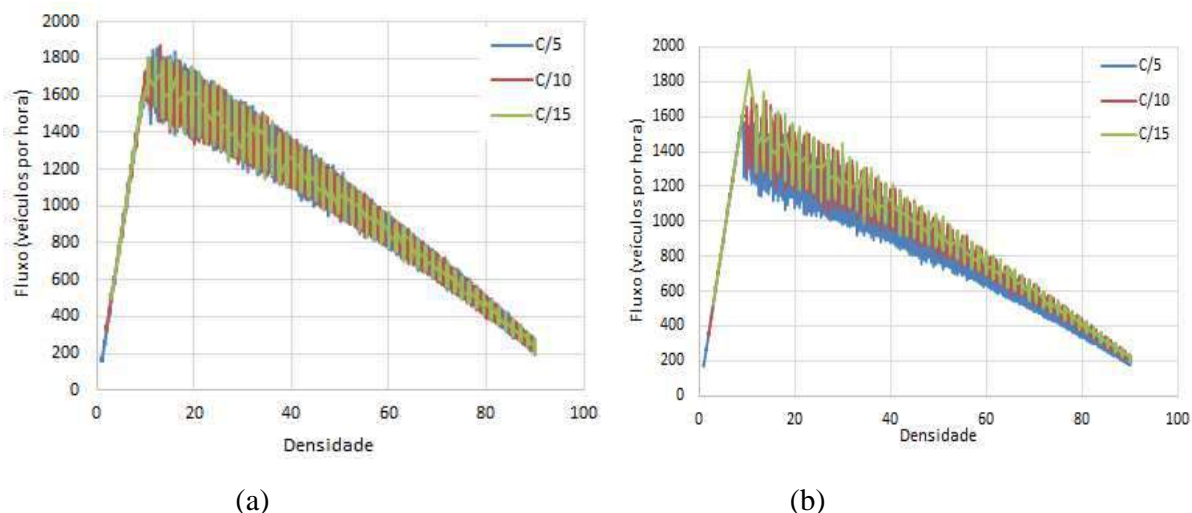


Figura 3: Diagramas fluxo-densidade – influência da discretização da via para política de aceleração (a) agressiva; (b) cautelosa.

Também neste caso há uma maior influência quando se utiliza uma política mais cautelosa de aceleração. Para melhor entendimento apresentamos, na Figura 4, o diagrama fluxo-densidade para o perfil mais cauteloso, mostrando, para cada valor da densidade, o valor médio do fluxo. Nessa figura também foi incluída a discretização $C/20$. Podemos observar que os resultados vão convergindo para uma mesma relação fluxo-densidade quando a discretização da via vai sendo refinada, mostrando uma convergência nos resultados.

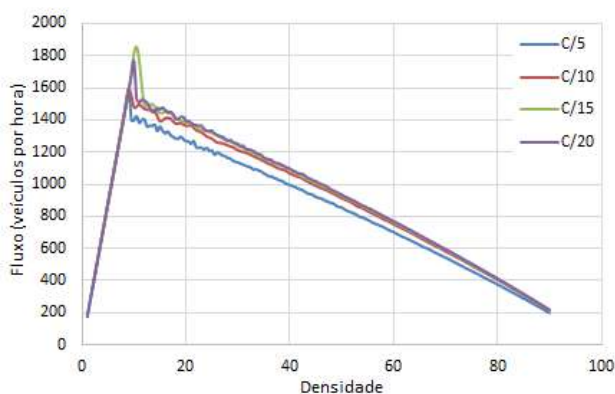


Figura 4: Diagrama fluxo-densidade – influência da discretização da via para a política de aceleração cautelosa (fluxo médio).

5- Conclusões

Este trabalho utiliza um modelo de NaSch modificado, que permite que sejam definidas diferentes políticas de aceleração/desaceleração. Verificou-se que essas diferentes políticas afetam a dinâmica do tráfego e devem ser consideradas quando se quer fazer uma análise mais precisa para simular o tráfego real. Verifica-se também que na região de fluxo livre, políticas de direção mais cautelosas conduzem a velocidades médias da via maiores. A situação se inverte quando começa a haver interação entre veículos, uma vez que os mais cautelosos demoram a retomar a velocidade cada vez que precisam diminuir sua velocidade devido a algum obstáculo, por exemplo, um veículo mais lento à sua frente.

Em relação à variância, resultados iniciais mostram que elas afetam progressivamente os resultados, e isso não se dá de forma uniforme para todos os perfis. Os resultados até o momento encontrados mostram que ela influencia mais naqueles perfis que representam uma política mais cautelosa ao acelerar/desacelerar. Esse estudo deve ter continuidade para se ter melhor entendimento desse funcionamento, e avaliar a forma mais conveniente para representar as diversas políticas de aceleração e sua influência em todas as situações já analisadas.

Os resultados que avaliam a influência da discretização da via mostram que todos eles convergem para uma mesma relação de fluxo-densidade, embora essa aparente convergência seja mais lenta para perfis mais cautelosos.

Além disso, os resultados confirmam a importância de se modelar, em todas as situações, a influência das diferentes políticas de aceleração na modelagem do tráfego viário, um dos principais objetivos do presente trabalho.

Bibliografia

1. Nagel, K., Schreckenberg, M.: A cellular automaton model for freeway traffic. *Journal de physique I* 2(12), 2221-2229 (1992).
2. Zamith, M, Leal-Toledo, R. C. P., Clua, E., Toledo, E. M., e Magalhes, G. V.: A new stochastic cellular automata model for traffic flow simulation with drivers behavior prediction. *Journal of computational science* 9, pp. 51-56 (2015).
3. Leal-Toledo, R. C. P., Magalhães, G.V. P., Miranda, I., Toledo, E. M., Traffic Flow Simulation Under the Influence of Different Acceleration Policies, *Int'l Conf. Scientific Computing (CSC'17)*, pp 90-94 (2017)
4. I. M. Almeida , R. C. P. Leal-Toledo, E. M. Toledo, D. C. Cacao, e G. V. P. Magalhães, Drivers' behavior effects in the occurrence of dangerous situations which may lead to accidents, aceito para ser apresentado no 5th Workshop on Traffic and Cellular Automata (a ser publicado *Lecture Note in Computer Science*), 2018.
5. Disponível em: < <http://traffyclab.ce.gatech.edu/node/66> > Acesso em 18/07/2018.
6. Disponível em <<https://www.telegraph.co.uk/news/science/science-news/8011029/Aggressive-and-timid-drivers-cause-traffic-jams-scientists-discover.html>> Acesso em 18/07/2018.
7. Zamith, M.P., Um Modelo de Autômato Celular Aplicado ao Tráfego Viário com Múltiplos Perfis de Condutores, Tese Doutorado, Pós-Graduação em Computação, Instituto de Computação-UFF.
8. Casella, G.; Robert, C. P.; Wells, M. T. [Generalized Accept-Reject sampling schemes](#). *Lectures Notes – Monograph Series*, pp. 342–347 (2004).

1. Dados gerais

TÍTULO: MÉTODOS NUMÉRICOS PARA O ESCOAMENTOS BIFÁSICO EM MEIOS POROSOS HETEROGÊNEOS EM AMBIENTES COMPUTACIONAIS DE ARQUITETURA de Memória HÍBRIDA.

Bolsista: Weber Guilherme Dias Ribeiro

Orientadora: Carla Osthoff Barros

Co-orientador: Frederico Luís Cabral

Programa Institucional de Bolsas de Iniciação Científica – PIBIC

Esse relatório cobre as atividades realizadas no período de 10 de Maio de 2018 a 20 de julho de 2018

2. Objetivos

O objetivo do presente trabalho é desenvolver conhecimento na área de processamento de alto desempenho para avaliar e otimizar uma aplicação de computação científica em máquinas de arquitetura paralela híbrida, com o auxílio das ferramentas de análise de código.

3. Introdução

A proposta desse projeto de iniciação científica é dar continuidade ao projeto de otimização do código “MÉTODOS NUMÉRICOS PARA O ESCOAMENTO BIFÁSICO EM MEIOS POROSOS HETEROGÊNEOS”, desenvolvido pelo grupo de pesquisa em Óleo e Gás do LNCC, em ambientes computacionais de arquitetura composta por processadores multicore interligados em um ambiente de memória distribuída.

Alguns pré-requisitos de conhecimentos são necessários antes do desenvolvimento da paralelização em um ambiente multicore, múltiplos nós de processamento e ferramentas de análise de desempenho, Vtune Amplifier, Intel Advisor e Intel Trace Analyser and Collector, sejam utilizadas.

Até o momento as atividades desenvolvidas tratam do estudo sobre elementos fundamentais que são a base do trabalho, portanto foram estudados e trabalhados temas

como os conceitos, básicos e avançados, das arquiteturas multicore e de processamento massivamente paralelo utilizando a arquitetura Intel haswell/broadwell/skylake.

Na seção 5, Resultados e Discussão, serão apresentadas as informações extraídas sobre os temas: pipeline, superscalar, arquiteturas Pentium 4 e Knights Corner.

4. Material e Métodos ou Metodologia

Para essa etapa inicial do projeto de pesquisa, o material e a metodologia ficaram restritos ao estudo de livros e artigos. Foram selecionados capítulos específicos dentro de livros didáticos comumente utilizados em disciplinas relacionadas à arquitetura e a organização de computadores, bem como artigos que se relacionam parcial ou totalmente com o conteúdo que foi estudado nessa etapa ainda primitiva do projeto.

Foram realizadas sínteses acerca de todo o material estudado, com o objetivo de eficiência didática e consulta futura.

Os livros e artigos utilizados para esses estudos iniciais foram escolhidos com base na sua correlação direta ou suficientemente parcial com o tema, e, também, no caso dos livros, por sua linguagem didática que fornece um entendimento mais genérico do tema.

Todos os materiais utilizados, livros e artigos, estão citados na seção 7, referências bibliográficas.

5. Resultados e Discussão

O passo inicial dos estudos foi entender os conceitos e posteriormente o funcionamento de uma pipeline e em seguida de uma arquitetura superescalar.

Referente à pipeline, alguns conceitos destacam-se por possuir correlação direta com a proposta do trabalho, como por exemplo as noções de speedup com base na conclusão de instruções por ciclo de relógio (sempre dependendo da quantidade de estágios da pipeline) e a questão dos pipeline hazards, que impedem uma execução contínua da pipeline, reduzindo o speedup em relação a um processador sem qualquer tipo de estratégia de paralelismo. Dentro dos principais pipeline hazards estudados, os que receberam mais atenção foram os data hazards (quando há conflito no acesso de um operando), e os control hazards (causados por previsões erradas acerca de branches).

Como exemplo prático, foi utilizada a pipeline do Intel 80486, que possui um tamanho relativamente conveniente para fins didáticos. com apenas 5 estágios : fetch, decode 1, decode 2, execute e write back.

Quanto ao tema de arquiteturas superescalares, foi dada atenção especial à questão do paralelismo a nível de instrução e as condições que impedem esse paralelismo, como dependência de dados (Read after Write, Write after Read e Write after Write),

dependência de condicionais/saltos(não há como prever se o programa irá executar o salto ou não) e o conflito de recursos. Políticas de emissão e conclusão de instruções como in-order issue/in-order completion, in-order issue/out-of-order completion, e out-of-order issue/out-of-order completion, também foram objeto de atenção especial.

Após a leitura desses temas iniciais, foi feito um breve estudo sobre a avaliação de desempenho de uma CPU genérica, com o intuito de aprender a fazer uma estimativa matemática do custo computacional de um conjunto de instruções, com o uso de grandezas simples, como o número de instruções, o tempo de ciclo e a quantidade de ciclos gasto por um tipo de instrução específica. Ainda nesse estudo, foi possível aprender a calcular taxas como MIPS(Million Instructions Per Second), e MFLOPS(Million Floating-point Operations Per Second). Esse estudo foi concluído com a dedução dos cálculos de desempenho que envolvessem um sistema com uma hierarquia de memória, e como o tempo gasto para processar um fluxo de instruções pode variar de acordo com a presença de múltiplas caches e diferentes frequências de cache hit e cache miss.

Antes de especificar um estudo sobre o Pentium 4 e o KNC, foi feita uma leitura na organização e funcionamento mais geral da arquitetura x86 da Intel. Esse estudo foi dirigido aos pontos da arquitetura x86 que mais se relacionavam com os conceitos de superescalar e pipeline. Analisar a classificação de registradores do tipo x86, bem como suas respectivas funções foi o passo inicial para começar a entender o básico dessa arquitetura. De forma objetiva, é possível separar os registradores x86 nas seguintes classes: general(de propósito geral), segment(que fazem referência a algum segmento de memória), flags(guardam condition codes e mode bits) e instruction pointer(armazena o endereço da instrução atual), além dos registradores específicos da FPU, mais precisamente, numeric(armazenam um número em ponto flutuante), control(controlam operações em ponto flutuante), status(refletem o status atual da FPU), e tag word(contém bits que informam a natureza do valor armazenado em dado registrador do tipo numeric). Em seguida foi feito um aprofundamento na constituição e funcionamento de duas classes de registradores, um pouco mais complexos: os do tipo flag e do tipo control. A última parte desse olhar inicial para a arquitetura x86, foi focada em entender o processo básico para se processar uma interrupção ou exceção nesse tipo de processador, bem como a natureza desses processos e suas classificações.

O primeiro exemplo prático tomado foi o Pentium 4. Novamente o foco do estudo estava relacionado com o ciclo de instruções, mais precisamente, a parcela do ciclo de instruções que se dá quase inteiramente no processador. As leituras tiveram como objetivo entender com mais detalhes o esquema de funcionamento básico de um Pentium 4:

- 1) O processador faz fetch de instruções na ordem do programa
- 2) Cada instrução é então traduzida em uma ou mais instruções do tipo RISC, conhecidas como micro-instruções.
- 3) O processador executa essas micro-instruções numa arquitetura superescalar.
- 4) O processador então envia os resultados das micro-instruções para os registradores envolvidos na instrução que compunha o fluxo original do programa.

Um ponto muito interessante e que vale a menção, é que a pipeline do Pentium 4 possui um tamanho consideravelmente grande: 20 estágios. Esse número pode ser ainda maior,

devido ao fato de algumas micro-instruções precisarem de múltiplos estágios de execução.

Foram estudados as partes mais relevantes desse processo geral descrito acima em 4 etapas, são eles:

Geração de micro-instruções: estágio em que uma unidade de fetch/decode busca instruções presente no fluxo original do programa e as traduz em uma, duas, três ou quatro micro-instruções(nesse caso, instruções do tipo RISC, de 118bits), armazenadas numa cache L1 denominada trace cache.

Ponteiro da próxima instrução da trace cache: lida com a seleção da próxima instrução da trace cache, e possui um sistema de branch prediction particular, que envolve a utilização de consultas de histórico num buffer especial, denominado Branch Target Buffer, afim de fazer previsões baseadas em experiências anteriores com aquela mesma entrada.

Fetch na trace cache: a trace cache recebe as micro-instruções já decodificadas e as armazena de forma sequencial, nesse estágio, essas micro-instruções são recuperadas de forma sequencial.

Drive : Esse é o quinto estágio da pipeline do Pentium 4, que entrega as instruções já decodificadas presentes na trace cache, para o módulo de renomeação/alocação.

Alocação: Esse estágio aloca recursos necessários para a execução. Tarefas de destaque incluem a criação de entradas no reorder buffer(ROB) e na filas de micro-instruções para o scheduler.

Renomeação de registradores: estágio que remapea os 8 registradores de uso geral e os 8 registradores de ponto flutuante em um outro conjunto de 128 registradores físicos.

Enfileiramento de micro-instruções: as micro-operações são então enviadas a uma das duas filas de espera possíveis, para aguardar espaço livre nos schedulers. Uma das filas é destinada a operações que envolvem acesso à memória, e a outra a operações que não acessam a memória.

Escalonamento(scheduling) e Despacho(dispatching): aqui, cada um dos schedulers busca na fila por micro-instruções que estejam preparadas para execução imediata(tenham seus operandos prontos e com recursos requeridos disponíveis), e então as despacha para a unidade de execução.

Já o Knights Corner, é uma implementação específica da arquitetura MIC(Many Integrated Core), que começou com o Larrabee em 2007, inicialmente projetada para ser um produto com características de GPU/HPC, mas logo em seguida reduzida somente a HPC, e apelidado de Knights Ferry (KNF). Nesse contexto, o KNC surge como um sucessor lógico do KNF.

O KNC é um objeto de estudo interessante, uma vez que seus cores são baseados na especificação P54C do Pentium original, obviamente com adesão de funcionalidades e estruturas, como o suporte 64 bits, melhorias no sistema de cache, um bus bidirecional especial, e outras, afim de garantir um desempenho condizente com o contexto temporal em que foi lançado.

6. Conclusão

Os estudos realizados são pré-requisito e formam a base de conhecimentos necessários para o desenvolvimento teórico sólido e servem de fundamentação para continuação no campo teórico, com a análise e comparação de outras arquiteturas Intel.

Ignorar os estudos básicos e iniciais, visando uma iniciação mais rápida em temas mais avançados, poderia tanto atrapalhar a compreensão futura, bem como empobrecer análises de resultados e conclusões posteriores

7. Referências Bibliográficas

STALLINGS, William. Arquitetura e organização de computadores. 8ª ed. São Paulo: Pearson Prentice Hall, 2010.

TANENBAUM, Andrew S.; ZUCCHI, Wagner Luiz. Organização estruturada de computadores. 5ª ed. São Paulo: Pearson Prentice Hall, 2007.

RAUBER, Thomas; RÜNGER, Gudula. Parallel programming: For multicore and cluster systems. Springer Science & Business Media, 2010.

“An Overview of Programming for Intel® Xeon® processors and Intel® XeonPhi™ coprocessors, James Reinders, Intel, 2012”,

“how-to-tune-applications-using-a-topdown-characterization”, Intel Developer Zone, 2004 HOW Series “Deep Dive”: Webinars on Performance Optimization – 2017 Edition (<https://colfaxresearch.com/how-series/>).

Using OpenMP – The Next Step – by Ruud van der Pas, Eric Stotzer and Christian Terboven (2017) An Introduction to Parallel Programming with OpenMP, PThreads and MPI – by Robert Cook

Relatório de Atividades: Inicialização de Fluidos para Animação Computacional

Bolsista: Allan Carlos Amaral Ribeiro Orientador: Gilson Antonio Giraldi

Tipo da Bolsa: BIBITI-LNCC

Período do Relatório: 01/08/2017 a 31/07/2018

17 de julho de 2018

1 Objetivos

Este projeto tem como objetivo estudar e desenvolver novos métodos para visualização científica de campos tensoriais, abordando a animação de fluidos. Além disso espera-se aplicar os conceitos de topologia de campos tensoriais e métodos de reação e difusão para iniciação de fluidos através da geração de campos tensoriais a partir de sketches (desenhos do usuário com 'rascunho' do campo).

Outro objetivo é o estudo do método GVF (Gradient Vector Flow) [12], sua aplicação para geração do campo tensorial a partir do sketch, e suas implicações na topologia do campo resultante, culminando na elaboração de artigos científicos descrevendo os resultados obtidos.

2 Introdução

Métodos para animação de fluidos são geralmente difíceis de se lidar dada a natureza inerentemente complexa da mecânica de fluidos. Atingir resultados satisfatórios mesmo relaxando algumas propriedades físicas é uma tarefa difícil. Nestas aplicações, a simulação da dinâmica de fluidos frequentemente recai na necessidade de conter ou controlar o movimento do fluido (geralmente artificialmente) sem perder a plausibilidade física. Assim, o fluido deve se comportar naturalmente, porém sem se mover de forma completamente livre para atender as necessidades da animação. Desta forma, o animador pode predefinir um caminho pelo qual o fluido deva seguir ou definir uma região para concentrá-lo, produzindo objetos fluidos, por exemplo.

O trabalho de [10] foi um dos primeiros a abordar o problema de controle de fluidos, usando um conjunto de imagens de frames definidos pelo usuário para guiar a simulação. Em [4], o fluido é controlado através de uma função de distância, que é usada para gerar uma força externa. Enquanto esses trabalhos focam o estado final do fluido, há os que buscam controlá-lo durante toda a simulação. A maioria usa forças externas para restringir caminhos pelo qual o fluido escoar. Em [2], uma força de controle é projetada de o campo de velocidade converge para o fluxo especificado. Outros métodos, como [6] e [8], são baseados na deformação da malha da simulação. Interpolação de campos é outra alternativa, tendo sido empregada em [7] de forma a gerar animações de fluidos incompressíveis.

Nesse projeto, temos o interesse em achar meios naturais de controlar fluidos, tanto em relação ao caminho, quanto ao estado final. Para isso, foram utilizados campos tensoriais como parte inerente do meio. Foi empregada uma abordagem baseada em malha, na qual cada voxel contém não apenas as quantidades típicas de fluido, mas também um tensor. A ideia principal é que os tensores em cada voxel

deverão desviar o fluido, alterando os vetores de velocidades. Essa abordagem apresentada em [3], usa tensores simétricos semidefinidos de segunda ordem, geralmente associados às estimativas de covariância, frequentemente encontrados em trabalhos relacionados à tractografia [11]. O método é uma adaptação de [9], em que o tensor é inserido na formulação matemática de forma que ele altera localmente o momento do fluido. Cada uma das etapas básicas da simulação (advecção, difusão e projeção) foram alteradas para levar em consideração a influência do tensor, refletindo as equações adaptadas.

No período desse relatório, desenvolvemos uma versão melhorada do método de projeção proposto em [3], com normalização adequada, de forma a evitar problemas numéricos encontrados em trabalhos anteriores. Também foram propostas diretrizes para a criação de campos tensoriais capazes de gerar simulações de fluido controladas. Foi avaliado como diferentes parâmetros envolvidos na construção dos campos podem afetar o comportamento do fluido, provendo uma visão de como tensores planares podem ser usados para conter o fluido e criar uma separação entre a forma desejada e seus arredores.

Houve duas contribuições nessa primeira etapa. Em primeiro lugar, foi proposta uma normalização que torna factível a projeção anisotrópica de campos tensoriais. Em segundo lugar, é proposta a partição de campos tensoriais em regiões bem definidas, visando principalmente permitir o fluxo de fluido no interior da região de interesse e inibir o fluxo na região externa.

3 Material e Métodos ou Metodologia

Nessa etapa do projeto adaptamos o método proposto por [3] para simulação estável de fluidos usando campo tensoriais definidos em malha regular. Uma grande vantagem deste método é sua estabilidade numérica devido sua formulação. Todos os métodos relatados a seguir foram implementados e aprimorados pelo bolsista.

O tensor é um modulador local de momento para o fluido que se move ao longo do campo, modelado pela seguinte equação diferencial:

$$\frac{\partial \mathbf{u}}{\partial t} = \mathbf{T}\mathbf{u} - \mathbf{u}, \quad (1)$$

em que \mathbf{u} é a velocidade e \mathbf{T} , um tensor simétrico positivo de segunda ordem. Esta é uma equação diferencial homogênea com coeficientes constantes no tempo, ou seja $\partial \mathbf{T} / \partial t = 0$. O campo tensorial B é estacionário, isto é, ele não varia com o tempo. A ideia é acoplar essa EDP às equações de Navier-Stokes de forma a reger o comportamento do fluido. Há várias formas de fazer tal combinação [3]. Para animação, uma maneira direta é usa o seguinte sistema modificado de equações para descrever a dinâmica dos fluidos.

$$\left\{ \begin{array}{l} \frac{\partial \mathbf{u}}{\partial t} = -(\mathbf{u} \cdot \nabla) \mathbf{u} - \frac{1}{\rho} \nabla + \nu \nabla \cdot (\mathbf{T} \nabla \mathbf{u}) + \mathbf{f} \\ \frac{\partial \mathbf{u}}{\partial t} = \mathbf{T}\mathbf{u} - \mathbf{u} \\ \nabla \cdot \mathbf{u} = 0 \\ \frac{\partial \rho}{\partial t} = -(\mathbf{u} \cdot \nabla) \rho + \theta \nabla \cdot (\mathbf{T} \nabla \rho) + s \\ \mathbf{u}(0) = \mathbf{u}_0, \end{array} \right. \quad \text{(puramente advectivo)} \quad (2)$$

onde ρ é a densidade do fluido, \mathbf{f} é o campo de forças externo, ν é a viscosidade, θ é o coeficiente de difusão e s um termo de fonte.

A equação (1) entra no modelo acima como uma equação puramente advectiva, trabalhando como forte restrição ao movimento do fluido. A Equação (1) deve, considerando o campo tensorial associado, restringir a movimentação do fluido ao longo de certos caminhos e defletir-lo em direção a caminhos desejados. Os tensores também afetam o termo difusivo das velocidades, já que ele multiplica o gradiente

da quantidade de fluido sendo difundida. Isso resulta em uma difusão não mais isotrópica. Os termos de viscosidade ν e a difusão θ afetam a norma do tensor \mathbf{T} em seus respectivos termos.

Na equação da densidade, foi adicionado o termo de fonte s , para inserir fluido no sistema. Já que o fluido que sair da região controlada pelos tensores dificilmente retornará, propõe-se desenvolver campos tensoriais que dissipem o momento do fluido nestas regiões. Essa é uma alternativa plausível ao fator de dissipação global geralmente utilizado por modelos de fluidos.

A advecção e a difusão foram implementadas como descrito em [3]. A melhoria proposta está no passo de projeção (apresentado a seguir) que visa tornar o campo de velocidades conservativo.

3.1 Método proposto

A etapa de projeção consiste na solução de um sistema de Poisson que mantém o campo de velocidades livre de divergente, um pré-requisito para a incompressibilidade do fluido. Mesmo que fumaça tenha sido o fluido escolhido para todas as simulações neste trabalho, pode-se considerar fumaça como um fluido incompressível obtendo resultados interessantes para a animação [1]. Observamos que a projeção original para regularizar a pressão acaba suavizando a influência dos tensores provida pela advecção e pela projeção. Deste modo, faz sentido que a pressão também esteja sujeita à influência do campo tensorial, que neste projeto é interpretado como uma propriedade do meio. A solução do sistema de Poisson é mais complexa, já que uma nova restrição foi adicionada ao sistema, resultando na seguinte equação:

$$\nabla \cdot (\hat{\mathbf{T}}\nabla p) = \frac{1}{\Delta t} \nabla \cdot \mathbf{u}, \quad (3)$$

onde $\hat{\mathbf{T}}$ é a versão normalizada do tensor original. Uma vez que a etapa de projeção é apenas um passo para garantir a incompressibilidade, não faz sentido ter nenhum tipo de ganho ou perda de momento. A questão aqui é qual normalização utilizar para que o sistema convirja.

Qualquer tensor de segunda ordem semidefinido em \mathbb{R}^3 (com autovalores $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$) pode ser decomposto em três componentes básicos: 1) linear: $\lambda_1 > 0, \lambda_2 = \lambda_3 = 0$; 2) planar: $\lambda_1 > 0, \lambda_2 = \lambda_3 = 0$; 3) isotrópico: $\lambda_1 > 0, \lambda_2 = \lambda_3 = \lambda_1$. Tensores lineares tendem a transportar o fluido em uma única direção. É o tipo mais desafiador para resolver numericamente, já que a troca de fluido pode ocorrer apenas em uma linha reta. Já que dois de seus autovalores são nulos, isso pode gerar descontinuidades locais. Tensores planares tendem a transportar o fluido no plano gerado pelos seus dois principais autovetores. Eles impõem restrições muito menores na troca de fluido do que os tensores lineares. No entanto, eles ainda apresentam um autovalor nulo que ainda pode gerar descontinuidades locais. Note que autovalores menores que um inibem a troca de fluido através da diminuição do momento do fluxo. Aqueles maiores que um tendem a aumentar as velocidades.

A criação de campos tensoriais deve levar em consideração todos esses aspectos para se obter o efeito de fluxo desejado. Nessa primeira etapa do projeto, tem-se o interesse em utilizar apenas um campo de tensores discretizado $B = \mathbf{T}_i$ para a simulação do modelo dado pela Equação 2, onde $1 \leq i \leq n_b$ e n_b é o número de tensores dentro do espaço representado por B . !!Frase confusa: Ele governa as etapas de advecção e a difusão que a estabilidade permite tensores arbitrários em qualquer lugar, ou seja, qualquer combinação de autovetores e autovalores!! Este campo governará as etapas de advecção e difusão, etapas cuja estabilidade permite tensores arbitrários em qualquer lugar, ou seja, qualquer combinação de autovetores e autovalores. Fluxos suaves são obtidos com distribuições suaves de autovalores com autovetores alinhados.

O passo de projeção anisotrópica, no entanto, é muito mais complicado. Cada tensor impõe localmente uma maior restrição a solução da Equação de Poisson. Se ele tiver autovalores nulos, trocas de fluido não poderão ocorrer em algumas direções, o que pode levar, em alguns casos, a um sistema sem solução. Tensores lineares e planares são um desafio nesse contexto. A inabilidade de calcular a pressão

resultaria em um campo de velocidades que não pode ser transformado em um campo livre de divergente. A referência [1] recomenda que a advecção seja realizada sempre em um campo de velocidade livre de divergente, para evitar geração de artefatos.

Propomos então garantir que autovalores não nulos satisfaçam $\lambda_i \geq 1$ na etapa de projeção, ou seja eles não reduzem as trocas de pressão. Isso é feito computando

$$\hat{\mathbf{T}}_i = \frac{\mathbf{T}_i}{\lambda_l}, \quad (4)$$

em que λ_l é o menor elemento não nulo de $\mathbf{T}_i \in B$. Isto é necessário até pra projeção de regiões isotrópicas e implica que não se pode ter tensores nulos. Note que o campo normalizado $\hat{B} = \hat{\mathbf{T}}_i$ só é necessário para a etapa de projeção.

É possível que um campo tensorial resulte em uma simulação instável e até que cause a divergência da projeção, isto é, se forem definidos autovalores muito altos. Além disso, descontinuidades no campo tensorial podem afetar negativamente a projeção anisotrópica. Assim, sugere-se um campo tensorial o mais suave possível e com autovalores moderados para permitir o cálculo adequado da pressão.

3.2 Construção de campos tensoriais 3D

Essa fase do projeto também focou na construção de campos tensoriais 3D $\hat{B} = \hat{\mathbf{T}}_i$ que tendam a manter o fluido no interior da região de interesse. Como o modelo proposto vem de uma formulação contínua, o desafio consiste em prover um campo tensorial que: 1) permita, de preferência, o fluido dentro da região de interesse; 2) inibe ou reduz o momento do fluido fora da região de interesse; 3) é suficientemente suave para evitar erros numéricos ao usar a projeção anisotrópica proposta (Equação 4).

O objetivo não é confinar o fluido dentro da área de interesse, isso iria requerer condições de contorno fora do escopo deste trabalho. Ao invés disso, a troca de fluido é permitida entre todas as regiões 3D. Considerando o número de graus de liberdade em três dimensões, a tarefa de confinar o fluxo de fluido sem condições de borda especiais não é fácil. Com uma malha de resolução suficientemente alta, no entanto, o método proposto permite de prover um campo tensorial capaz de confinar quase que perfeitamente o fluido em uma região 3D.

É proposto o uso de camadas de tensores planares sobre a superfície da região de interesse, formando uma casca. A normal da camada (ou seja, o autovetor associado a λ_3) deve ser alinhado às normais da superfície da região de interesse. O objetivo da casca é reduzir a troca de fluidos ente as regiões internas e externas.

Formalmente, a proposta é dividir o campo em três regiões mutuamente exclusivas

$$B = S \cup I \cup O, \quad (5)$$

onde S é o conjunto de tensores formando a casca, I forma a região interna, $S \cup I$ é a região de interesse e O é o conjunto de tensores externos a essa região, tendo $S \cap I = \emptyset$, $S \cap O = \emptyset$ e $I \cap O = \emptyset$. A casca é formada por tensores planares

$$S = \{\mathbf{P}_i, \quad 1 \leq i \leq n_S, \quad \mathbf{P}_i = \lambda_S \cdot (\mathbf{I} - \mathbf{n}_i \cdot \mathbf{n}_i^T)\}, \quad (6)$$

onde n_S é o número de tensores em S , \mathbf{n}_i é o vetor unitário normal à superfície no voxel i , e λ_S é a amplificação desejada para todos os tensores de S . Nos experimentos, \mathbf{n}_i foi definido como a normal no ponto mais próximo da superfície do objeto. Essa abordagem direta exigiria uma grade de alta resolução para representar corretamente superfícies de alta curvatura. Mas pode ser simplesmente adaptada para qualquer aplicação. Note que $\lambda_1 = \lambda_2 = \lambda_S$ para todos os tensores em S . Se λ_S for menor que um, a casca tenderá a anular as velocidades sobre ela. Se maior que um, ela adicionará momento às velocidades

tangentes aos tensores planares. Ambas as possibilidades podem ser exploradas para gerar efeitos no entorno da região I .

Os conjuntos I e O representam regiões cujo objetivo é fornecer meio para o fluxo de fluido. Faz sentido que tais regiões sejam preenchidas com tensores isotrópicos na forma

$$I = \{\mathbf{S}_i, \quad 1 \leq i \leq n_I, \quad \mathbf{S}_i = \lambda_I \cdot \mathbf{I}\}, \quad (7)$$

$$O = \{\mathbf{O}_i, \quad 1 \leq i \leq n_O, \quad \mathbf{O}_i = \lambda_O \cdot \mathbf{I}\}, \quad (8)$$

em que n_I e n_O são as quantidades de tensores nos conjuntos I e O , respectivamente. Note que $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_I$ para todos os tensores de I , e $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_O$ para todos os tensores de O . Os autovalores em cada região devem ser ajustados de acordo com o efeito desejado para o fluxo. Se igual a um, o fluido se comporta da mesma forma que o método clássico isotrópico. Se menor que um o campo tende a reduzir as velocidades. Se maior que um, o fluido tende a ganhar momento. Nos experimentos realizados, evitou-se utilizar valores maiores que um. Isso pode trazer problemas de estabilidade devido ao constante aumento da velocidade. Para se diminuir a velocidade e tornar as regiões de interesse mais visíveis, podemos usar o campo tensorial externo O com autovalores menores que um. O resultado é a grande redução das velocidades e densidades fora da casca. Em suma, foram definidos $\lambda_I = 1$ e $\lambda_O < 1$ se a região interna permite o fluxo de fluido. Para simular fluxo preferencialmente na casca, pode-se definir $\lambda_I < 1$.

4 Resultados e Discussão

Para avaliar a validade das propostas desta etapa do trabalho, foram desenvolvidos diferentes tipos de campos tensoriais. Como de costume em simulações de fluido, cada um dos parâmetros teve de ser escolhido cuidadosamente para se obter o efeito desejado.

Primeiramente foram utilizados campos tensoriais sintéticos como uma região esférica isotrópica envolta por uma casca planar. Os testes consistiam em aplicar diferentes tipos de forças a partir do centro da esfera e observar o comportamento do fluido na região. Tais testes mostraram que, com um ajuste fino de parâmetros, tensores planares podem ser utilizados tanto para conter o fluxo de fluido dentro da região de interesse como podem ser utilizados para espalhar o fluido pela superfície. Os experimentos também mostraram que um mesmo campo tensorial poderia ser utilizado para se atingir diferentes efeitos na animação do fluido.

Em seguida experimentos mais elaborados foram realizados através de campos tensoriais obtidos para malhas poligonais triangulares. Forças externas foram aplicadas na região interior à malha, usando tanto a propriedade de contenção quando a de espalhamento das cascas de tensores planares para espalhar o fluido pela região interior da malha.

Em vista dos resultados, o método proposto se mostrou competente em conter o fluido, em sua maior parte, na região interna à malha, sendo necessário apenas o uso de malhas mais refinadas para que regiões pequenas e superfícies com alta curvatura pudessem ser capturadas pelo campo tensorial. Os resultados foram bons o suficiente para viabilizar a submissão de um artigo completo para a conferência SIBGRAPI 2018, que é B1 em Ciência da Computação [5].

5 Conclusões

A etapa atual do projeto se mostrou de fundamental importância para o entendimento de como tensores, especialmente as cascas de tensores planares, podem ser utilizados para controlar uma simulação de fluido

e de sua utilidade para definir campos tensoriais que sejam uteis para inicialização de fluido. Tal estudo será de grande valia ao abordar métodos de reação-difusão para geração de campos tensoriais.

Além disso, o método de normalização proposto tornou possível a utilização da etapa de projeção anisotrópica para uma maior variedade de campos tensoriais, provando ser fundamental para a utilização em campos tensoriais arbitrários resultados da aplicação de métodos de GVF sobre sketches de usuários.

Numericamente, as simulações se mostraram estáveis mesmo com campos tensoriais arbitrários. Visualmente foi possível verificar claramente o benefício da projeção proposta na contenção e controle do fluido. Mais especificamente, a partição do espaço 3D em três regiões em conjunto com a projeção anisotrópica gerou resultados satisfatórios. Os resultados desta fase do projeto foram submetidos ao congresso SIBGRAPI 2018 [5].

Referências

- [1] Robert Bridson. *Fluid simulation for computer graphics*. AK Peters/CRC Press, 2015.
- [2] Yootai Kim, Raghu Machiraju, and David Thompson. Path-based control of smoke simulations. In *Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 33–42. Eurographics Association, 2006.
- [3] Marcelo Caniato Renhe. *A Stable Tensor-Based Deflection Model for Controlled Fluid Simulations*. dissertation, COPPE - Universidade Federal do Rio de Janeiro, 2017.
- [4] Marcelo Caniato Renhe, Antonio Oliveira, Cláudio Esperança, and Ricardo Marroquim. Enhanced target driven smoke morphing. In *Graphics, Patterns and Images (SIBGRAPI), 2012 25th SIBGRAPI Conference on*, pages 213–220. IEEE, 2012.
- [5] Allan Amaral Ribeiro, Marcelo Caniato Renhe, Marcelo Bernardes Vieira, Gilson Giraldi, and Cláudio Esperança. Anisotropic fluid modeling and simulation using tensor fields. In *31th SIBGRAPI, Foz do Iguaçu, Brazil, October 29 - November 01, 2018*, 2018 (Submitted).
- [6] Syuhei Sato, Yoshinori Dobashi, Kei Iwasaki, Tsuyoshi Yamamoto, and Tomoyuki Nishita. Deformation of 2d flow fields using stream functions. In *SIGGRAPH Asia 2014 Technical Briefs*, page 4. ACM, 2014.
- [7] Syuhei Sato, Yoshinori Dobashi, and Tomoyuki Nishita. A combining method of fluid animations by interpolating flow fields. In *SIGGRAPH ASIA 2016 Technical Briefs*, page 4. ACM, 2016.
- [8] Syuhei Sato, Yoshinori Dobashi, Yonghao Yue, Kei Iwasaki, and Tomoyuki Nishita. Incompressibility-preserving deformation for fluid flows using vector potentials. *The Visual Computer*, 31(6-8):959–965, 2015.
- [9] Jos Stam. Stable fluids. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 121–128. ACM Press/Addison-Wesley Publishing Co., 1999.
- [10] Adrien Treuille, Antoine McNamara, Zoran Popović, and Jos Stam. Keyframe control of smoke simulations. *ACM Transactions on Graphics (TOG)*, 22(3):716–723, 2003.
- [11] Anna Vilanova, Shangjian Zhang, Gordon Kindlmann, and D Laidlaw. An introduction to visualization of diffusion tensor imaging and its applications. In *Visualization and Processing of Tensor Fields*, pages 121–153. Springer, 2006.
- [12] Chenyang Xu and Jerry L Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on image processing*, 7(3):359–369, 1998.

1. Dados Gerais

Aplicação das ferramentas Intel Parallel Studio para modernização de código para métodos numéricos de diferenças finitas para solução de equações diferenciais parciais em arquitetura Intel Haswell/Broadwell.

Bolsista: Gabriel Pinheiro da Costa

Orientadora: Carla Osthoff

Co-orientador: Frederico Luís Cabral

Bolsa de Iniciação Tecnológica pelo CNPq

Esse relatório cobre as atividades realizadas entre agosto de 2017 até julho de 2018

2. Objetivos

O objetivo é o desenvolvimento do conhecimento na área de processamento de alto desempenho para avaliar e otimizar o desempenho de uma aplicação de computação científica em uma máquina de arquitetura paralela haswell/broadwell, com o uso das ferramentas de análise de código

3. Introdução

O trabalho almeja realizar um estudo das oportunidades de otimização de código para a solução de equações diferenciais parciais através das ferramentas do Intel Parallel Studio.

Os métodos numéricos destinados à solução dessas equações são contemplados pela possibilidade vasta do uso de ferramentas e técnicas cada vez mais efetivas, que buscam aumentar o desempenho dessas aplicações, reduzindo o tempo de execução e permitindo entradas de dados de maior magnitude.

No contexto de otimização, ganha forte destaque a técnica de paralelização, que permite que os dados sejam processados simultaneamente por mais de uma unidade de processamento, sem comprometer a qualidade do resultado final, e reduzindo o tempo necessário para chegar até esse resultado.

As ferramentas do Intel Parallel Studio trabalham com informações do comportamento das instruções do código durante a compilação e a execução. A partir dessa gama de informações é possível extrair conclusões acerca da aplicação e seu desempenho atual, e portanto nortear um plano de otimização para essa aplicação, buscando um aproveitamento ótimo do hardware utilizado.

4. Material e Métodos ou Metodologia

Para a realização dos testes das diferentes implementações do método HOPMOC foram usadas uma máquina contendo um acelerador Intel Xeon Phi KnightsCorner (KNC) com 60 cores de 1.053 GHz e 4 threads por core, e uma máquina contendo um acelerador Intel Xeon Phi KnightsLanding (KNL) com 68 cores de 1.4 GHz e 4 threads por core.

5. Resultados e Discussão

Talvez a forma mais básica para realizar a paralelização do HOPMOC seja a inserção de pragmas OpenMP nos loops que realizam maior número de iterações, ou seja, o loop de interpolação linear e os loops que calculam os operados explícitos e operadores implícitos. Essa versão inicial foi nomeada de Naive.

```

1 #pragma simd
2 #pragma omp for
3 for (int i = head+1 ; i <= N-2 ; i+=2) {
4     ANNOTATE_ITERATION_TASK(loop_HOP_EXP_1);
5     U_old[i] = alfa*(U_new[i-1] + U_new[i+1]) +
6         (1 - 2*alfa)*U_new[i];
7 }
8 #pragma omp single
9     head = (head+1)%2;
10
11 #pragma simd
12
13 #pragma omp for
14 for (int i = head+1; i <= N-2 ; i+=2) {
15     ANNOTATE_ITERATION_TASK(loop_HOP_IMP_1);
16     U_old[i] = (U_new[i] + alfa*U_old[i-1] + alfa
17         *U_old[i+1]) / (1+2*alfa);
18 }
19 #pragma omp single
20     head = (head+1)%2;
21
22 #pragma simd
23 #pragma omp for
24 for (int i = head+1 ; i <= N-2 ; i+=2) {
25     ANNOTATE_ITERATION_TASK(loop_HOP_EXP_2);
26     U_new[i] = alfa*(U_old[i-1] + U_old[i+1]) +
27         (1 - 2*alfa)*U_old[i];
28 }
29 #pragma omp single
30     head = (head+1)%2;
31
32 #pragma simd
33 #pragma omp for
34 for (int i = head+1; i <= N-2 ; i+=2) {
35     ANNOTATE_ITERATION_TASK(loop_HOP_IMP_2);
36     U_new[i] = (U_old[i] + alfa*U_new[i-1] + alfa
37         *U_new[i+1]) / (1+2*alfa);

```

As análises feitas pelo Intel Thread Advisor para essa estratégia revelam um ganho de desempenho com paralelismo extremamente limitado, devido à baixa quantidade de instruções por loop. O Thread Advisor também sugere, para um possível ganho de desempenho, mudanças estruturais no código que possibilitem um agrupamento das instruções dentro dos loops. Assim deve-se aumentar a quantidade de instruções por loop e reduzir a quantidade total de loops.

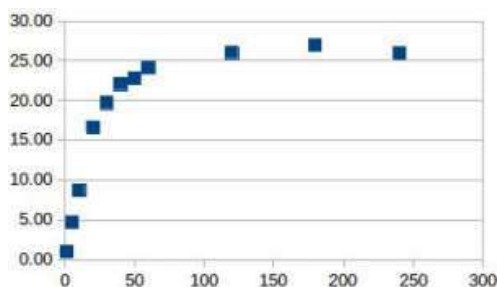
Dessa necessidade surge um novo código, onde as instruções responsáveis pelo cálculo dos operadores explícitos e implícitos são agrupadas em loops únicos, sem comprometer o resultado final. Essa nova versão foi apelidada de Chunk, ou Cluster Points.

```

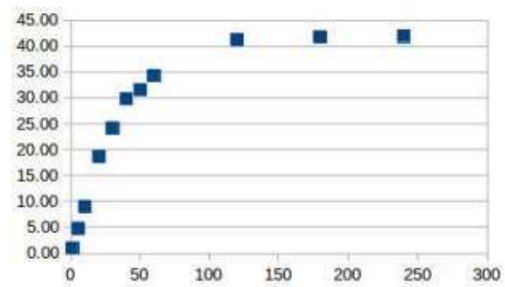
1 #pragma omp for
2 for (int i = 4 ; i <= N-3 ; i+=4) {
3     U_til[i-1] = alfa*(U_til_til[i-2] +
4         U_til_til[i]) + (1 - 2*alfa)*
5         U_til_til[i-1];
6     U_til[i+1] = alfa*(U_til_til[i] +
7         U_til_til[i+2]) + (1 - 2*alfa)*
8         U_til_til[i+1];
9     U_til[i] = (U_til_til[i] + alfa*U_til[i
10        -1] + alfa*U_til[i+1])/(1+2*alfa);
11    U_new[i] = alfa*(U_til[i-1] + U_til[i
12        +1]) + (1 - 2*alfa)*U_til[i];
13 }
14 #pragma omp for
15 for (int i = 2 ; i <= N-5 ; i+=4) {
16     U_til[i] = (U_til_til[i] + alfa*U_new[i
17        -1] + alfa*U_new[i+1])/(1+2*alfa);
18     U_new[i] = alfa*(U_til[i-1] + U_til[i
19        +1]) + (1 - 2*alfa)*U_til[i];
20     U_new[i-1] = (U_til[i-1] + alfa*U_new[i
21        -2] + alfa*U_new[i])/(1+2*alfa);
22     U_new[i+1] = (U_til[i+1] + alfa*U_new[i
23        ] + alfa*U_new[i+2])/(1+2*alfa);
24 }

```

Os gráficos subsequentes comparam os speedups por número de threads das versões Naive e Chunk em um acelerador KNC, comprovando o ganho da segunda abordagem.



(a) Naive



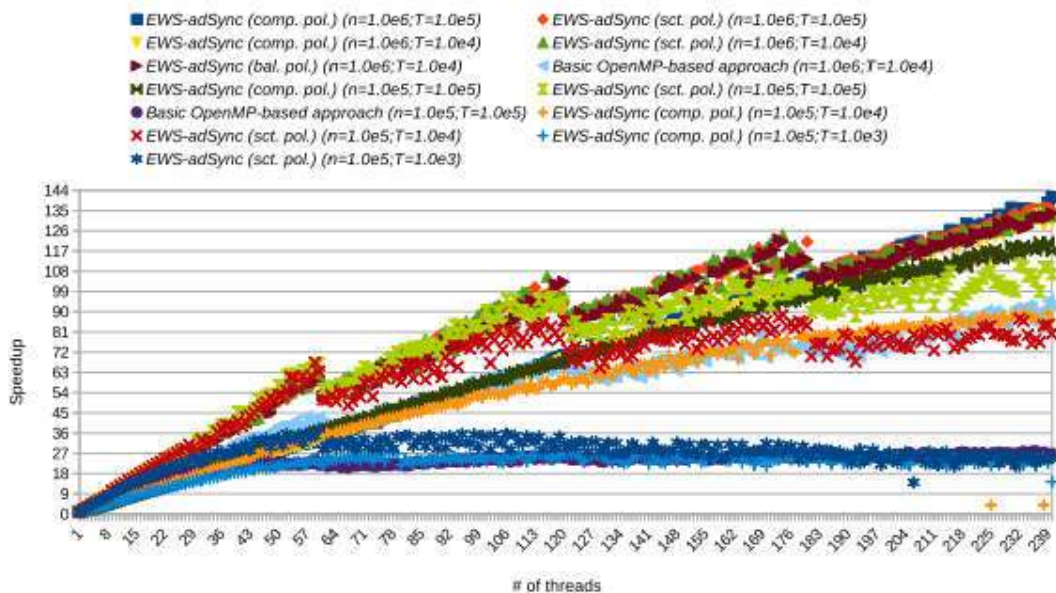
(b) Chunk

A tabela a seguir mostra os comparativos de tempo entre as duas versões, sendo executadas, também, em um Xeon Phi KNC.

Simulation Results		
Threads	Naive Version Time	Chunk Version Time
1	769.17	784.94
5	163.38	163.89
10	88.74	87.20
20	46.24	42.21
30	39.05	32.57
40	34.90	26.42
50	33.70	24.94
60	31.93	22.98
120	29.52	19.08
180	28.51	18.84
240	29.60	18.79

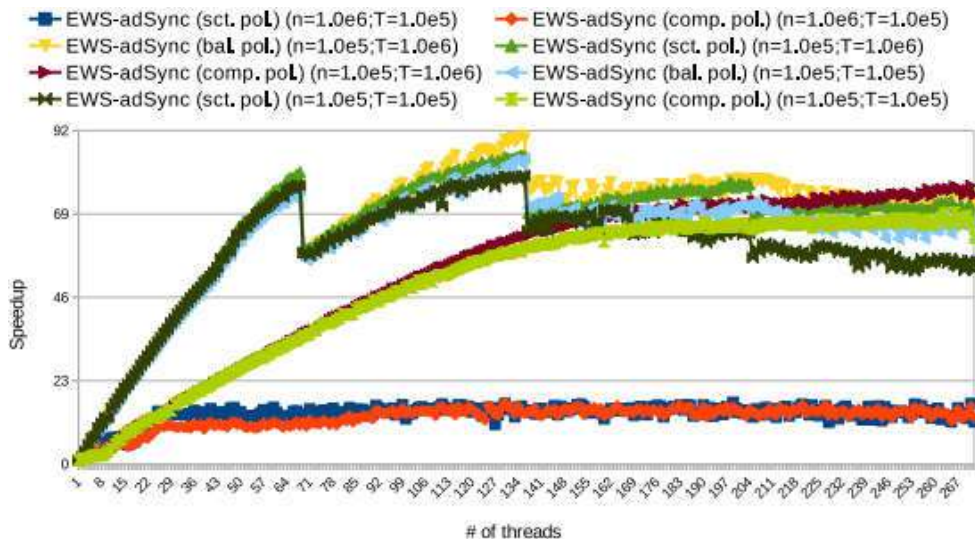
Outra estratégia desenvolvida para a otimização do método em arquitetura paralela foi a EWS-adSync. Essa implementação conta com uma distribuição manual das threads, evitando o tempo gasto em scheduling em tempo de execução (runtime). A malha com os pontos a serem calculados é particionada de forma explícita em partes fixamente divididas, e cada uma dessas partes é designada a uma thread, permanentemente, até o fim da execução. Essa atribuição manual e estática de partições por threads garante que só haverá um scheduling de threads durante toda a execução.

Essa implementação também minimiza a troca de mensagens entre as threads, e o tempo gasto em barreiras de sincronização. Na implementação EWS-adSync, uma thread precisa somente de informações de threads adjacentes a ela para dar sequência à execução. Dessa forma é aplicado um mecanismo de sincronização explícito, que consiste em um vetor booleano, em que cada entrada corresponde a uma thread, e o valor da entrada indica se os dados podem ou não podem ser acessados pelas threads vizinhas em um dado instante de tempo.



O gráfico acima mostra o speedup por número de threads para execuções feitas em um acelerador KNC, usando diferentes políticas de thread binding: balacend, compact e scatter. O speedup com o uso da estratégia EWS juntamente com a sincronização manual de threads vizinhas, com o uso de uma política compact usando 240 threads chega a 140x, como mostra a imagem.

Testes semelhantes realizados na KNL apresentaram os seguintes resultados



Aqui, o speedup máximo da estratégia EWS com sincronização manual de threads adjacentes é de 96x, com a política de thread binding scatter.

6. Conclusões

Os testes realizados em arquiteturas multicore paralelas (KNC e KNL) apontam para ganhos consideráveis de desempenho em ambas as implementações desenvolvidas. Tanto a estratégia de agrupamento de pontos (Cluster Points) quanto a de divisão explícita de threads juntamente com sincronização manual (EWS-adSync) mostraram-se soluções possíveis para os problemas encontrados. As ferramentas de análise da Intel se provaram extremamente eficientes em apontar comportamentos que resultam em perda de desempenho e em nortear a busca por alterações no código original afim de obter um desempenho ótimo.

No caso da estratégia Cluster Points, a junção de pequenas tarefas em uma grande tarefa (iteração do loop) possibilitou um aumento de speedup, e na abordagem EWS-adSync, esse ganho de speedup foi possível pela distribuição manual das threads (uma divisão explícita de trabalho) que reduziu o tempo de scheduling ao tornar desnecessário a divisão implícita de threads em tempo de execução, e também pela sincronização somente entre threads vizinhas, que eliminou o tempo gasto em barreiras de sincronização.

7.Referências

RAUBER, Thomas; RÜNGER, Gudula. **Parallel programming: For multicore and cluster systems**. Springer Science & Business Media, 2010.

Cabral, F.L.; Osthoff, C.; Costa, G.P.; Gonzaga de Oliveira, S.L.; Brandão, D.; Kischinhevsky, M. *An OpenMP Implementation of the TVD Hopmoc Method Based on a Synchronization Mechanism Using Locks Between Adjacent Threads on Xeon Phi (TM) Accelerators*. Lecture Notes in Computer Science. Springer International Publishing, 2018, v. 3, p. 701-707. [doi: 10.1007/978-3-319-93713-7_67](https://doi.org/10.1007/978-3-319-93713-7_67)

Cabral F.L.; Osthoff C.; Costa G.; Brandão D.; Oliveira S.L.G.; Kischinhevsky M. *Tuning up TVD HOPMOC method on Intel MIC Xeon Phi Architectures with Intel Parallel Studio Tools Workshop on Applications for Multicore Architecture (WAMCA 2017) - International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD 2017) - Campinas - São Paulo - Brasil - Outubro de 2017*.

TEIXEIRA, Thiago Daniel Quimas Simões. Estratégias de otimização para um Método Numérico para resolução de problemas de Convecção-Difusão. 2016. 72 f. Trabalho de Conclusão de Curso(Graduação em Sistema de Informação) - Universidade Estácio de Sá, Petrópolis, 2016.

RELATÓRIO PARCIAL

1) Dados gerais

Título: Técnicas de Aprendizagem de Máquina na Segmentação de Imagens

Médicas de Ultrassom Intravascular

Bolsista: Jefferson da Silva Fernandes de Azevedo

Orientador: Pablo Javier Blanco

Local: LNCC/MCTIC, Petrópolis

Período: 20-06-2018 a 20-07-2018

2) Objetivos

Os objetivos realizados e em processo de realização conforme o plano de trabalho são os seguintes

1. Pesquisar sobre os métodos de aprendizagem de máquina baseados em redes neurais profundas
2. Pesquisar sobre o problema da segmentação de imagens de ultrassom intravascular

3) Introdução

O processamento de imagens médicas é um tópico central na caracterização e no entendimento do início e progresso de doenças cardiovasculares. Por exemplo, através da identificação de estruturas anatômicas nas paredes das artérias coronárias, é possível estabelecer correlações com o risco de desenvolvimento de eventos cardiovasculares. As imagens de tipo IVUS (ultrassom intravascular) são adquiridas por meio de um procedimento de cateterismo. Neste procedimento, um sensor girando a altíssima velocidade é inserido nas artérias coronárias de um paciente até uma região de interesse. A partir do início do estudo, o sensor é retirado a uma velocidade controlada e pequena em relação com a velocidade de rotação do mesmo. Desta forma, o sensor é capaz de captar informações de “seções transversais” do vaso ao longo do seu percurso de retirada da artéria. A área de pesquisa na segmentação de imagens IVUS é extremamente ativa desde os anos 90, e o problema continua ainda em aberto. O desafio apresentado por estas imagens consiste nas altas taxas de ruído, cuja característica é a de ser multiplicativo, que dificulta o processo de segmentação automática. neste projeto abordaremos o problema da segmentação automática de imagens de IVUS. Isto

implica em segmentar as estruturas anatômicas de interesse que se podem observar nos vasos coronários, tais como lúmen, membrana elástica externa, e diferentes tipos de tecidos constituintes da parede arterial. Para abordar este problema utilizaremos métodos baseados no aprendizado de máquina, em particular redes neurais de tipo convolucionais, as quais têm demonstrado maior capacidade para cumprir tarefas de segmentação e classificação de imagens em problemas de visão computacional.

4) Metodologia

O projeto se apoia na utilização de redes neurais profundas convolucionais. Este tipo de arquitetura tem demonstrado capacidade de lidar com problemas de visão computacional, possibilitando abordar diversos tipos de questões tais como segmentação, classificação, etiquetamento, dentre outros. Em particular, utilizaremos dados de imagens de ultrassom intravascular disponíveis no grupo HeMoLab de forma a treinar as redes neurais.

5) Resultados e Discussão

Atualmente temos realizado um estudo preliminar sobre imagens de tipo IVUS assim como também um estudo preliminar sobre a metodologia de aprendizagem de máquina baseada em redes neurais profundas.

Além disso, realizaram-se atividades visando o pré-processamento dos dados disponíveis de imagens IVUS com vistas a sua utilização no treinamento de redes neurais.

Dado o estágio inicial do projeto, não há resultados específicos a serem discutidos com maior detalhe.

6) Conclusão

As atividades de pesquisa e desenvolvimento começaram há um mês. As mesmas estão de acordo com o planejamento proposto no plano de trabalho.

Assinatura:

Petrópolis, 15 de Julho de 2018

LABORATÓRIO NACIONAL DE COMPUTAÇÃO CIENTÍFICA

RELATÓRIO DE ATIVIDADES

Título do Projeto: *Processamento e Visualização de Imagens de Microscopia Confocal*

Nome: **Matheus Alves de Moura**

Orientador: Gilson Antonio Giraldi

Tipo de bolsa: PIBIT

Período: 01/08/2017 – 20/07/2018

Rio de Janeiro

Julho – 2017

Objetivos

1. Segmentação de imagens de microscopia confocal do ovário de *Drosophila*.
 - 1.1. Implementação da técnica de *Graph cut*;
 - 1.2. Implementação da técnica de *Convex Shape Prior*;
 - 1.3. Implementação da técnica de *Firefly*;
 - 1.4. Implementação de método de *region growing*.
2. Fazer reconstrução tridimensional de imagens *z stack*.

1.Introdução

Por meados de 721 a.C. há os primeiros relatos de que lentes obtidas de um cristal duma rocha foram utilizadas para aumentar objetos de interesse do homem. Desde então, com o desenvolvimento de diversas tecnologias, como o desenvolvimento de lâmpadas mais potentes e o entendimento dos lasers, alguns tipos de microscópios surgem no meio científico, como o microscópio de força atômica, o microscópio eletrônico de varredura, o microscópio confocal dentre outros. Este último patenteado por volta de 1956 funciona compartilhando parte da física óptica de um microscópio de fluorescência, porém sua principal diferença é que o microscópio confocal, a partir dos *pins holes* que cortam grande parte da luz desfocada, possibilita recriar imagens 3D do espécime de interesse, podendo assim, desvendar diversos mistérios da biologia[1,2].

Uma imagem de microscopia confocal de um único plano focal tem fundo preto e pontos claros são onde o anticorpo ou fluoróforo ficaram excitados e emitiram fótons em determinado comprimento de onda. Técnicas para melhoramento dessas imagens surgem, com o intuito de extrair informações ou gerar modelos geométricos das estruturas de interesse [3].

Para o processamento das imagens, uma técnica que é usada e que mostra resultados significativos [6], é a técnica de corte de grafos para segmentação de imagens. Esta técnica trata a imagem com um grafo, o que permite que a imagem possa ser separada entre fundo e objeto. Neste trabalho, as técnicas aqui citadas foram aplicadas à imagens de *Drosophila*, mostradas a frente.

3.Materiais e Métodos

3.1 Graph Cut

Assim como outras técnicas [3], a segmentação de imagens a partir de corte de grafos tem como objetivo separar os pixels em “objeto” (*object*) e “fundo” (*background*). Esta técnica tem como fundamento uma equação de energia computada sobre um grafo.

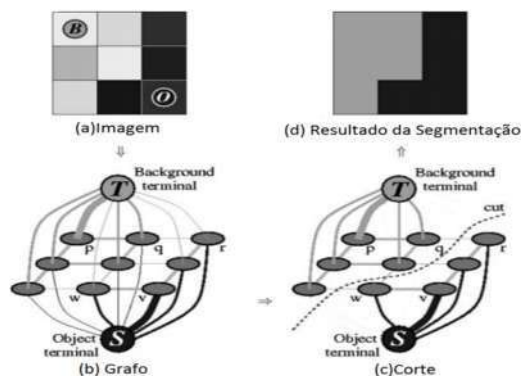


Figura 1. Representação de uma imagem 3x3 e seu corte por grafos. (a) Os conjuntos B e O representam pontos de fundo e objeto (sementes), respectivamente. (b) Representação do grafo inicial. (c) Ilustrações de corte de grafos. A linha tracejada indica o corte do grafo, realizada pela técnica que é discutida neste trabalho. (d) representa o resultado final de uma imagem segmentada. (Adaptado de [4]).

Um grafo pode ser definido por um conjunto de pontos ligados por arestas, onde cada aresta pode ter um peso diferente. Na técnica de segmentação de imagens discutida em [4], para cada pixel há um nó pertencente a um conjunto V que está ligado por arestas de um conjunto E , obtidas em relação a uma determinada vizinhança. Assim, uma imagem qualquer pode ser representada por um grafo $G = (V, E)$, onde V são os nós do grafos (pixels na imagem) e E , as ligações (arestas) entre os nós. Note que, assim como na figura 1, existem ainda dois nós terminais, o S que representa o objeto, e o T que representam o *background*.

3.2 Convex shape prior

Como discutido em [6], a técnica de *Convex shape prior* é utilizada para separar dois objetos conexos. Como mostra a figura 2, a técnica é baseada no cálculo de pontos em curvas na borda da região conexa. Há a ligação do ponto mais fundo de todas as curvas da borda (ou o SCP, do inglês *steepest concave point*) ao ponto mais fundo da curva mais próxima (ou o NBP – *nearest boundary point*). A linha azul da figura 2 mostra a rasterização entre os pontos supracitados.

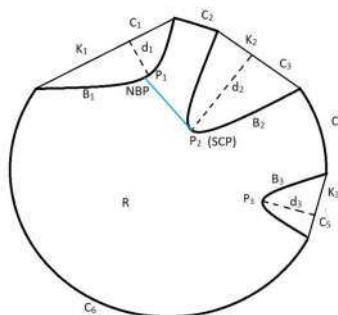


Figura 2. Região conexa R para segmentação. C_2, C_4 e C_6 são as curvas da borda de R , ao passo que C_1, C_3 e C_5 são chamadas cordas (cavidades na borda do objeto). K_1, K_2 e K_3 são as retas que unem as bordas das cordas. B_1, B_2, B_3 são curvas limites com os mesmos pontos finais que as retas K_1, K_2, K_3 . P_1, P_2 e P_3 são os pontos com as maiores distâncias d_1, d_2 e d_3 da corda K respectiva a sua curva C . P_2 é ponto com maior d , logo é o SCP. O ponto P_1 é o ponto mais próximo de P_2 sendo assim o NBP. Em azul, a rasterização entre os pontos SCP e NBP, como previsto pelo método. Adaptado de [6].

3.2.2 Convex shape prior adaptado

Os passos adicionados nesta adaptação é que, ao passo que a rasterização é feita, os pontos mais profundos (os pontos P nas imagens) que não foram utilizados na etapa anterior (na imagem 2 o exemplo é P_3), são utilizado agora como *input* da mesma técnica. Ou seja, as curvas B que foram segmentadas pelo método deixam de fazer parte da imagem e um novo corte é realizado na nova imagem.

3.3 Firefly

A técnica discutida em [7] é uma meta-heurística chamada de *firefly*, ou vaga-lume, que é uma técnica baseada no comportamento de busca destes animais após sua bioluminescência. Este modelo meta-heurístico já foi utilizado antes no processamento e segmentação de imagens, sendo assim uma técnica possivelmente viável para segmentar as imagens de interesse deste trabalho.

O algoritmo baseado em [7] num primeiro momento avalia o brilho da imagem, para que assim possa ser dito quais pixels são os vaga-lumes. Em seguida, é calculado a distância entre eles e assim são movidos em direção uns aos outros, de acordo com o brilho de cada um. Com este processo, é possível obter o *firefly* mais brilhante. Por fim, são geradas aleatoriamente novas soluções para o algoritmo.

Como o algoritmo tem parte randomizada, foi implementado junto a ela um método simples para que fosse possível haver a reprodutibilidade dos resultados. Uma única imagem passa pelo algoritmo dez vezes, ou seja, geram dez imagens como resultados, e o resultado final que o algoritmo mostra na realidade é a média aritmética destas dez imagens.

4. Resultados parciais

A Figura 3 contém o resultado de alguns planos de uma microscopia confocal do ovário de *Drosophila*. As imagens originais foram adquiridas pelo grupo de trabalho em um microscópio confocal de varredura a laser.

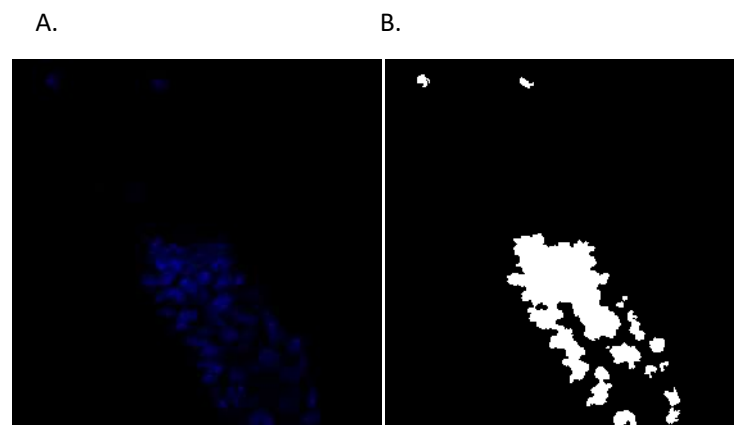


Figura 3. Aplicação da técnica 3.1 citada em [6]. A contém a imagem original e em B o resultado binarizado.

Como pode ser visto o resultado obtido não mostrou eficiência quanto à segmentação das células. O método aqui utilizado aglomerou as regiões que antes pareciam ser separadas, o que, a priori, inviabiliza a técnica. Com o intuito de melhorar os resultados da técnica de corte de grafos, a técnica de *Convex Shape Prior* (CSP) foi implementada.

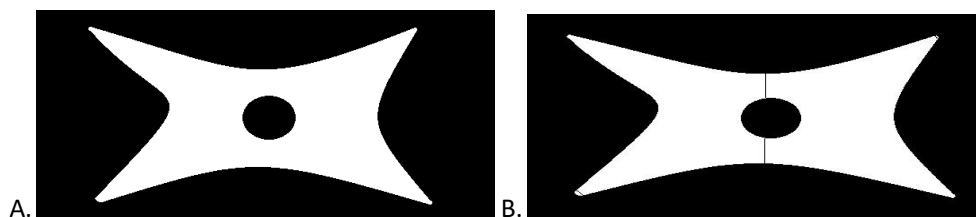


Figura 4. Implementação da técnica *Convex Shape Prior*. Em A, imagem teste e em B, resultado da técnica.

Em um primeiro momento, a figura 4 é uma figura teste que foi utilizada para testar a implementação da técnica CSP. Em um primeiro momento, a técnica se mostrou promissora, pois solucionava, na imagem teste, uma problemática, que era das células justapostas. Como os objetos a serem separados eram múltiplos, a técnica precisava de uma adaptação, que foi a técnica discutida no ponto 3.2.2, que foi otimizada para separar concavidades na borda do objeto a ser separado. A figura 5A é parte binarizada de uma das imagens a serem segmentadas. Como mostra a figura 5B, o resultado obtido mostrou-se incoerente com o esperado. Especula-se que esta técnica não funcionou na imagem original devido a sua alta complexidade, visto que a técnica CSP adaptada se baseia nas bordas do objeto e se essa for muito irregular, o resultado será diferente do esperado.

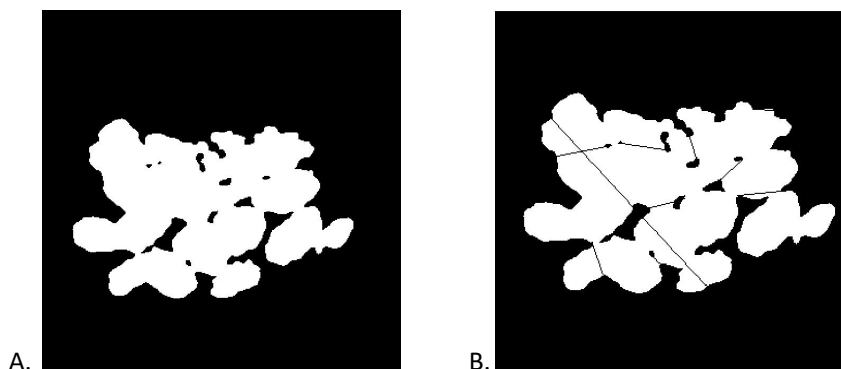


Figura 5. Implementação da técnica CSP adaptada. Em A, objeto conexo binarizado obtido a partir de um corte em uma das imagens a serem segmentadas. Em B, resultado da implementação *Convex Shape Prior adaptada*.

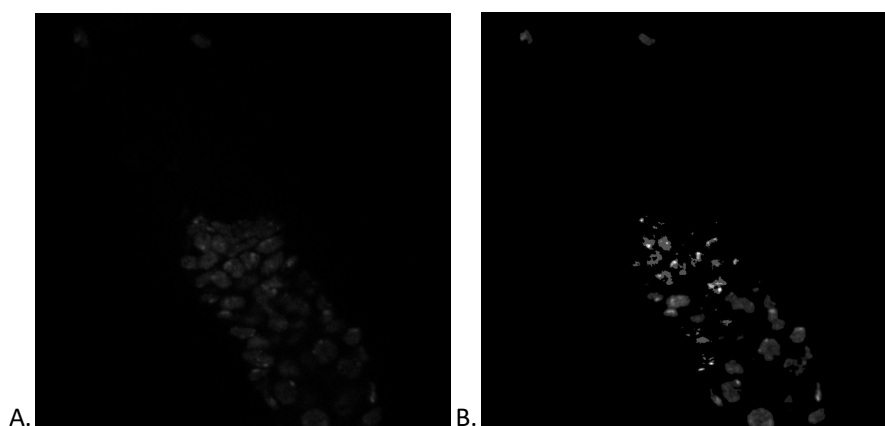


Figura 6. Implementação da técnica *firefly*. Em A, imagem a ser segmentada em tom de cinza. Em B, resultado da segmentação pelo método.

Após o insucesso das duas primeiras técnicas, uma revisão de alguns métodos já existentes de segmentação de imagens foi feita. Um dos métodos encontrados foi o método discutido no ponto 3.3 deste trabalho, que é a técnica meta-heurística de *firefly*. A figura 6B mostra o resultado desta técnica, onde é possível ver que os núcleos das células nos objetos conexos agora estão separados.

Com os centros de cada célula, um método de filtragem e crescimento de região precisa ser implementado para concluir o projeto e assim, ser possível segmentar as fatias ópticas do ovário de *Drosophila*.

5. Conclusão

- A segmentação de imagens é uma área que cresce e é imprescindível;
- Duas das três técnicas abordadas não demonstraram bons resultados;
- A técnica *firefly* conseguiu separar os centros dos objetos de interesse, possibilitando que, com o crescimento de região, os objetos de interesse possam ser segmentados.

Bibliografia

1. Wanderley De Souza W. (ed), 2010. **Microscopia Óptica: fundamentos e Aplicações às Ciências Biomédicas**. SBME, Rio de Janeiro.
2. David J. Stephens, Victoria J. Allan. **Light Microscopy Techniques for Live Cell Imaging**. Science, 2003:300, 5616, pp. 82-86. DOI: 10.1126/science.1082160.
3. L. Rosenblum, R. Earnshaw, J. L. Encarnação, H. Hagen, A. Kaufman, G. Klimenko, S. and Nielson, F. Post, and D. Thalmann. **Scientific visualization : advances and challenges**. New York Academic Press 1994, 1994.
4. Yuri Boykov, Gareth Funka-Lea, **Graph Cuts and Efficient N-D Image Segmentation**. International Journal of Computer Vision 70(2), 109–131, 2006.
5. James Jonkman and Claire M. Brown. **Any Way You Slice It—A Comparison of Confocal Microscopy Techniques**. *J Biomol Tech*. 2015 Jul; 26(2): 54–65.
6. Jin Qi, B. Wang, N. Pelaez, I. Rebay, R. W. Carthew, A. K. Katsaggelos, L. A. Nunes Amaral. **DROSOPHILA EYE NUCLEI SEGMENTATION BASED ON GRAPH CUT AND CONVEX SHAPE PRIOR**. ICIIP, 2013.
7. Paulo S. Rodrigues, Guilherme A. Wachs-Lopes, Horst R. Erdmann, Monael P. Ribeiro, Gilson A. Giraldi. **Improving a firefly meta-heuristic for multilevel image segmentation using Tsallis entropy**. Pattern Anal Applic DOI 10.1007/s10044-015-0450-x, 2015.